



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 111121

TO: Minh-Tam Davis
Location: cm1/8a01/8e12
Art Unit: 1642
Friday, January 02, 2004

Case Serial Number: 09/554945

From: Mary Jane Ruhl
Location: Biotech-Chem Library
CM1-6A06
Phone: 605-1155

maryjane.ruhl@uspto.gov

Search Notes

Examiner Davis,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
CM-1, Rm. 6-A-06
605-1155

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:20 ; Search time 46 Seconds

(without alignments)
1614.870 Million cell updates/sec

Title: ABU66614

Perfect score: 2399

Sequence: 1 MGFLGTGTLVLVLPQAF.....EKGILKEEAEAKRIYSSL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	2399	100.0	468	21 AAB18926	A novel polypeptid
2	2399	100.0	468	22 AAU12216	Human PRO5990 poly
3	2399	100.0	468	22 AAB87600	Human PRO5990. Ho
4	2399	100.0	468	23 ABG95925	Human secreted/tra
5	2399	100.0	468	24 ABU69097	Human PRO polypept
6	2399	100.0	468	24 ABU69120	Human PRO polypept
7	2399	100.0	468	24 ABU71580	Human secreted pol
8	2399	100.0	468	24 ABU72026	Novel human secret
9	2399	100.0	468	24 ABU72183	Human PRO polypept

10	2399	100.0	468	24 ABU66614	Human PRO polypept
11	2399	100.0	468	24 ABU66890	Human secreted/tra
12	2399	100.0	468	24 ABU59695	Novel secreted and
13	2399	100.0	468	24 ABG73287	Human neurosecret
14	2399	100.0	567	22 AA013905	Human polypeptide
15	2392	99.7	468	22 AAW93910	Human polypeptide,
16	2384	99.4	468	21 AAY81955	Human HagII prote
17	2271	94.7	470	22 ABG12976	Novel human diago
18	2102.5	87.6	471	23 ABB57046	Mouse ischaemic co
19	1974	82.3	396	24 AA016322	Human secreted pro
20	594	24.8	316	22 ABG14006	Novel human diago
21	475	19.8	212	22 ABG14003	Novel human diago
22	184	7.7	665	21 AAB18278	Plasmodium falcipa
23	180.5	7.5	1881	23 ABP73809	Candida albicans e
24	179	7.5	1558	21 AAB18324	Plasmodium falcipa
25	179	7.5	1786	18 AAW24790	P. falciparum live
26	179	7.5	1787	23 AAU96699	Plasmodium falcipa
27	178.5	7.4	1392	20 AAY06999	Resin protein seq
28	175.5	7.3	1427	12 AAR10534	Human 160kd mediat
29	174.5	7.3	1164	19 AAW40537	Group B streptococ
30	173	7.2	1099	19 AAW40538	Mutant C-beta prot
31	172.5	7.2	1164	17 AAR85781	Group B Streptococ
32	172.5	7.2	1164	21 AAY84459	Amino acid sequenc
33	172.5	7.2	1979	21 AAB18171	Plasmodium falcipa
34	171	7.1	1093	19 AAW40540	Mutant C-beta prot
35	170	7.1	1213	22 AAW40016	Human polypeptide
36	169.5	7.1	1788	22 ABB62995	Drosophila melanog
37	169	7.0	1129	21 AAY84462	Amino acid sequenc
38	168	7.0	1164	19 AAW40541	Mutant C-beta prot
39	168	7.0	1164	21 AAY84463	Amino acid sequenc
40	167	7.0	493	13 AAY26944	F.falciparum USA 9
41	166.5	6.9	1135	21 AAY84460	Amino acid sequenc
42	166	6.9	1128	19 AAW40539	Mutant C-beta prot
43	166	6.9	1128	21 AAY84461	Amino acid sequenc
44	166	6.9	1192	21 AAB18165	Plasmodium falcipa
45	166	6.9	2633	22 ABG06505	Novel human diago

ALIGNMENTS

RESULT 1

ID AAB18926 standard; Protein; 468 AA.

XX AAB18926;

XX AC

XX DT 08-FEB-2001 (first entry)

XX DE

XX DE A novel polypeptide designated PRO5990.

XX DE

KW Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;

KW PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;

KW PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;

KW PRO4424; PRO4422; PRO4430; PRO4499; PRO4599; tumour; obesity; diabetes;

KW insulinemia; kidney disorder; Bergers disease; nephropathy;

KW Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;

KW Crohns disease.

XX OS Homo sapiens.

XX FH

XX FT Key

FT Modified-site

FT Location/Qualifiers

FT /note= "N-glycosylation site"

FT /note= 68..71

FT /note= 70..73

FT /note= "casein kinase II phosphorylation site"

FT /note= 82..85

FT /note= "casein kinase II phosphorylation site"

FT /note= 97..100

FT /note= "casein kinase II phosphorylation site"

FT /note= 125..128

FT /note= "casein kinase II phosphorylation site"

FT /note= 143..148

FT Modified-site /note= "N-myristoylation site"
 FT 147...150
 FT /note= "casein kinase II phosphorylation site"
 FT 188...191
 FT /note= "casein kinase II phosphorylation site"
 FT 217...220
 FT /note= "casein kinase II phosphorylation site"
 FT 239...244
 FT /note= "N-myristoylation site"
 FT 265...268
 FT /note= "casein kinase II phosphorylation site"
 FT 289...292
 FT /note= "casein kinase II phosphorylation site"
 FT 305...308
 FT /note= "casein kinase II phosphorylation site"
 FT 320...323
 FT /note= "casein kinase II phosphorylation site"
 FT 326...329
 FT /note= "casein kinase II phosphorylation site"
 FT 346...349
 FT /note= "N-glycosylation site"
 FT 350...353
 FT /note= "N-glycosylation site"
 FT 362...365
 FT /note= "casein kinase II phosphorylation site"
 FT 369...372
 FT /note= "casein kinase II phosphorylation site"
 FT 382...385
 FT /note= "casein kinase II phosphorylation site"
 FT 386...389
 FT /note= "casein kinase II phosphorylation site"
 FT 387...390
 FT /note= "casein kinase II phosphorylation site"
 FT
 FN WO200056889-A2.
 XX
 XX
 PD 28-SEP-2000.
 XX
 PF 01-MAR-2000; 2000WO-US05601.
 XX
 PR 23-MAR-1999; 99US-0125774.
 PR 23-MAR-1999; 99US-0125778.
 PR 24-MAR-1999; 99US-0125826.
 PR 31-MAR-1999; 99US-0127035.
 PR 05-APR-1999; 99US-0127706.
 PR 21-APR-1999; 99US-0130359.
 PR 27-APR-1999; 99US-0131270.
 PR 27-APR-1999; 99US-0131272.
 PR 04-MAY-1999; 99US-0132371.
 PR 04-MAY-1999; 99US-0132379.
 PR 04-MAY-1999; 99US-0132383.
 PR 25-MAY-1999; 99US-0135750.
 PR 08-JUN-1999; 99US-0138166.
 PR 20-JUL-1999; 99US-0144791.
 PR 03-AUG-1999; 99US-0146970.
 PR 09-DEC-1999; 99US-0170262.
 XX
 XX (GETH) GENENTECH INC.
 XX
 PI Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
 PI Stewart TA, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2000-628263/60.
 DR N-PSDB; AAA96353.
 XX
 PT Novel secreted and transmembrane polypeptides useful for diagnosing
 PT tumour in a mammal, for identifying agonists and antagonists of the
 PT polypeptide and for therapeutic use
 XX
 PS Claim 12; Fig 36; 222pp; English.
 XX
 CC The present sequence represents a secreted or transmembrane polypeptide.

CC The specification describes polypeptides designated PRO1484, PRO4334,
 CC PRO1122, PRO1899, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
 CC PRO4356, PRO4352, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
 CC PRO6030, PRO4424, PRO4430 and PRO4499. PRO1889 polypeptide is
 CC useful for diagnosing tumour in a mammal. The polypeptides, their
 CC agonists and antagonists are useful treating a condition associated with
 CC expression or activity of the polypeptide. Conditions treated include
 CC obesity, diabetes or hyper- or hypo-insulinemia. The polypeptides are
 CC capable of inducing proliferation of mammalian kidney mesangial cells
 CC and are therefore useful for treating kidney disorders associated with
 CC decreased mesangial cell function such as Bergers disease or other
 CC nephropathies associated with Schonelein-Henoch purpura, celiac disease,
 CC dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
 CC to generate transgenic animals for use in development and screening of
 CC therapeutically useful reagents and also for chromosome identification
 CC and tissue typing.
 XX
 SQ Sequence 468 AA;
 Query Match 100.0%; Score 2399; DB 21; Length 468;
 Best Local Similarity 100.0%; Pred. No. 4.1e-146;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGFLGTGTLVLVLPPIQAFPPKPGSQSKLNRLSARPLNEQIAEAEDEKIKKTYPP 60
 DB 1 MGFLGTGTLVLVLPPIQAFPPKPGSQSKLNRLSARPLNEQIAEAEDEKIKKTYPP 60
 QY 61 ENKPGQSNYSFVDNLNLLKATEKIEKROSIKSSPLDNKLVNVEDVSTQKRLIDDY 120
 DB 61 ENKPGQSNYSFVDNLNLLKATEKIEKROSIKSSPLDNKLVNVEDVSTQKRLIDDY 120
 QY 121 DSTKGLDHKFODDPDGLHOLDGTLTAEDIVHKTAARIYEENDRAVFDKIVSKLNLGL 180
 DB 121 DSTKGLDHKFODDPDGLHOLDGTLTAEDIVHKTAARIYEENDRAVFDKIVSKLNLGL 180
 QY 181 ITESQAHTLEDEVAEVLQKLISKANNEYEDPNKPTSWTENQAGKIPKVTPTMAAIQDGL 240
 DB 181 ITESQAHTLEDEVAEVLQKLISKANNEYEDPNKPTSWTENQAGKIPKVTPTMAAIQDGL 240
 QY 241 AKGENDETVSNTLTITNGLERTKTYSEDNFEELQYFPNFYALLKSIDSEKAEKETLI 300
 DB 241 AKGENDETVSNTLTITNGLERTKTYSEDNFEELQYFPNFYALLKSIDSEKAEKETLI 300
 QY 301 TIMKTLLIDFVKMWKYGITISPEEGVSVYLENDEMIALQTKNLEKNATDNISKLPFAPSE 360
 DB 301 TIMKTLLIDFVKMWKYGITISPEEGVSVYLENDEMIALQTKNLEKNATDNISKLPFAPSE 360
 QY 361 KSHBETDSTKEBAAMEKEYSGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNIEWLK 420
 DB 361 KSHBETDSTKEBAAMEKEYSGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNIEWLK 420
 QY 421 KHKKGKEDYDLKMRDFFINKQADAYVEKGILDKKEAEATKRIYSSL 468
 DB 421 KHKKGKEDYDLKMRDFFINKQADAYVEKGILDKKEAEATKRIYSSL 468
 RESULT 2
 AAU12216
 ID AAU12216 standard; Protein; 468 AA.
 XX
 AC AAU12216;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human PRO5990 polypeptide sequence.
 XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.
 XX
 OS Homo sapiens.
 XX

CC polypeptide, H is a PRO20233 polypeptide and I is a PRO1890
CC polypeptide. The sample comprises a cell suspected of expressing the A,
CC B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with
CC a detectable label or is attached to a solid support. The proteins are
CC useful for linking a bioactive molecule to a cell expressing a
CC polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive
CC molecule is a toxin, a radiolabel or an antibody. The bioactive molecule
CC causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies
CC against them are useful for modulating a biological activity of a cell
CC expressing a polypeptide designated as A, B, C or D or E, F, G, H, or
CC I. The cell is killed. The proteins are useful for identifying
CC agonists or antagonists, for the preparation of a medicament useful in
CC the treatment of a condition which is responsive to the proteins, as
CC molecular weight markers for protein electrophoresis purposes, and as
CC therapeutic agents for treating sports-related joint problems,
CC articular cartilage defects, osteoarthritis or rheumatoid arthritis.
CC Nucleic acids encoding the proteins are useful as hybridisation probes,
CC in chromosome and gene mapping, in the generation of anti-sense RNA and
CC DNA, for the preparation of the proteins, to generate transgenic or
CC knockout animals which are useful in the development and screening of
CC therapeutic useful reagents, for chromosome identification, and in gene
CC therapy. The antibody is useful as a therapeutic agent, in a diagnostic
CC assay and for affinity purification of the protein from recombinant
CC cell culture natural sources. The present sequence represents a novel
CC secreted or transmembrane protein of the invention.

XX SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 23; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGFWILVLPQAPFKPGSQDKSLHRELNSAERPLNQIAEAEEDKIKTYPP 60
Db 1 MGFLGTGFWILVLPQAPFKPGSQDKSLHRELNSAERPLNQIAEAEEDKIKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLLKATTEKEKEKEKQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKATTEKEKEKEKQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120

Qy 121 DSTKSGLDHKEFDQDPDGLHQLDGTPLTAEDIVHKIARIYEENDRAVDFDKIVSKLNLGL 180
Db 121 DSTKSGLDHKEFDQDPDGLHQLDGTPLTAEDIVHKIARIYEENDRAVDFDKIVSKLNLGL 180

Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTTPMAAIQDGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTTPMAAIQDGL 240

Qy 241 AKGENDETQSVNTLITNGLERTKTYSEDNPEELQYFPNFVALLKSIDSEKEAKEKETLI 300
Db 241 AKGENDETQSVNTLITNGLERTKTYSEDNPEELQYFPNFVALLKSIDSEKEAKEKETLI 300

Qy 301 TIMKTLIDFVKQWKYGTISPEEGVSYLENDEMIALQTNKLEKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKQWKYGTISPEEGVSYLENDEMIALQTNKLEKNATDNISKLPAPSE 360

Qy 361 KSHETDSTKEEAARKEKEYGSLDKSDTNDNPNCGKTDPEKGTTEAYLEAIRKNIEWLK 420
Db 361 KSHETDSTKEEAARKEKEYGSLDKSDTNDNPNCGKTDPEKGTTEAYLEAIRKNIEWLK 420

Qy 421 KHKKKGNKEDYDLKMRDFINKQADAVVEKGIIDKEBAEAIKRIYSSL 468
Db 421 KHKKKGNKEDYDLKMRDFINKQADAVVEKGIIDKEBAEAIKRIYSSL 468

RESULT 5
ABU69097
ID ABU69097 standard; Protein; 468 AA.
XX AC ABU69097;
XX XX
XX 02-JUN-2003 (first entry)
XX

DE Human PRO polypeptide #18.
XX Human; PRO polypeptide; secreted and transmembrane protein; cancer;
KW non-insulin dependent diabetes mellitus; septic shock; stroke;
KW rheumatoid arthritis; graft-versus-host disease; cardiac ischaemia;
KW psoriasis; inflammatory bowel disease; asthma; anti-diabetic;
KW cytostatic; immunosuppressive; antirheumatic; antiarthritic;
KW cerebroprotective; vasotropic; antipsoriatic; antiinflammatory;
KW antiasthmatic.
XX Homo sapiens.
OS
XX US2003008348-A1.
PN
XX 09-JAN-2003.
PD
XX 26-DEC-2001; 2001US-0035855.
PF
XX 14-MAY-1999; 99WO-US10733.
PR 02-DEC-1999; 99WO-US28551.
PR 22-DEC-1999; 99WO-US30720.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 02-MAY-2000; 2000WO-US14042.
PR 02-JUN-2000; 2000WO-US15264.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 01-DEC-2000; 2000WO-US32678.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19892.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 15-MAY-1998; 98US-085579P.
PR 15-DEC-1998; 98US-112514P.
PR 22-DEC-1998; 98US-113300P.
PR 23-DEC-1998; 98US-113430P.
PR 23-DEC-1998; 98US-113605P.
PR 22-JAN-1999; 99US-116843P.
PR 23-MAR-1999; 99US-125774P.
PR 23-MAR-1999; 99US-125778P.
PR 24-MAR-1999; 99US-125826P.
PR 31-MAR-1999; 99US-127035P.
PR 03-APR-1999; 99US-127706P.
PR 13-APR-1999; 99US-129122P.
PR 21-APR-1999; 99US-130359P.
PR 25-MAY-1999; 99US-135750P.
PR 08-JUN-1999; 99US-138166P.
PR 20-JUL-1999; 99US-144791P.
PR 03-AUG-1999; 99US-146970P.
PR 29-OCT-1999; 99US-162506P.
PR 16-AUG-2001; 2001US-0931836.
XX (GETH) GENENTECH INC.
PA
XX Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
PI Stewart TA, Watanabe CK, Wood WI, Zhang Z;
PI WPI; 2003-341326/32.
DR N-PSDB; ACA06116.
DR
XX New PRO polypeptides and nucleic acid molecules, useful for diagnosing
PT or treating diabetes mellitus, cancers, septic shock, inflammatory
PT bowel disease or asthma, or in gene therapy, chromosome identification
PT or tissue typing -
XX
PS Claim 12; Fig 36; 196pp; English.
XX
CC The present invention relates to the isolation of novel human PRO
CC polypeptides, and the polynucleotide sequences encoding them. The
CC PRO polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides and polynucleotides are useful in diagnosing or
CC treating non-insulin dependent diabetes mellitus, cancers, septic
CC shock, rheumatoid arthritis, graft-versus-host disease, stroke, cardiac

ischaemia, psoriasis, inflammatory bowel disease or asthma. The PRO polynucleotide sequences may be used as hybridisation probes in chromosome and gene mapping, or in generating antisense RNA and DNA. They are also useful in preparing PRO polypeptides, in assays to identify other proteins or molecules involved in binding reaction, to generate transgenic animals or knockout animals, which in turn are useful in the development and screening of therapeutically useful reagents, for chromosome identification, and tissue typing. The PRO polypeptides and nucleic acid molecules are also useful in gene therapy, and as molecular weight markers for protein electrophoresis purposes. Anti-PRO antibodies may be used in diagnostic assays for PRO polypeptides, or for the affinity purification of the polypeptides from recombinant cell culture or natural sources. ABU69080-ABU69102 represent the human PRO polypeptides of the invention.

Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.le-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTWILVLPVLPQAPKPGSGQSKLHNRLSAERPLNQIAAEEDKIKKTYPP 60
Db 1 MGFLGTGTWILVLPVLPQAPKPGSGQSKLHNRLSAERPLNQIAAEEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLKKAITEKEKIEKERSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLKKAITEKEKIEKERSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy 121 DSTKSGLDHFKFDDPDGHLQDGLPTLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKFDDPDGHLQDGLPTLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180

Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIEKVTWMAAODGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIEKVTWMAAODGL 240

Qy 241 AKENDETVNTLTLTNGLERRTKTYSSEDNFEELQYFPNFYALLKSIDSEKAKEKETLI 300
Db 241 AKENDETVNTLTLTNGLERRTKTYSSEDNFEELQYFPNFYALLKSIDSEKAKEKETLI 300

Qy 301 TIMKTLIDFVKVMVKYGTISPESGVSYLENLDEMIALQTNKLEKANATDNISKLFPAPSE 360
Db 301 TIMKTLIDFVKVMVKYGTISPESGVSYLENLDEMIALQTNKLEKANATDNISKLFPAPSE 360

Qy 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDONSFGKTDPEKGTAYLEAIRKNIEWLK 420
Db 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDONSFGKTDPEKGTAYLEAIRKNIEWLK 420

Qy 421 KHDKKNKEDYDLSKMRDFINKQADAYVEKGILDKGEAEAIKRIYSSL 468
Db 421 KHDKKNKEDYDLSKMRDFINKQADAYVEKGILDKGEAEAIKRIYSSL 468

RESULT 6
ABU69120
ID ABU69120 standard; Protein; 468 AA.
XX
XX ABU69120;
XX
XX
DT 02-JUN-2003 (first entry)
XX
DE Human PRO polypeptide #18.
XX
XX Human; secreted and transmembrane protein; bone disorder; obesity;
KW cartilage disorder; sports injury; arthritis; diabetes mellitus;
KW hypo-insulinaemia; obesity; hyper-insulinaemia; thalassaemia;
KW haemoglobin-associated disorder; kidney disorder; Berger disease;
KW mesangial cell function; nephropathy; Schönlein-Henoch purpura;
KW celiac disease; dermatitis herpetiformis; Crohn's disease;
KW anorectic; antiarthritis; antidiabetic; antianaemic; nephrotropic;
KW antinflamatory.

OS	Homo sapiens.	
XX		
PN	US2003032061-A1.	
XX		
PD	13-FEB-2003.	
XX		
XX		
PF	26-DEC-2001; 2001US-0036214.	
XX		
PR	14-MAY-1999; 99WO-US10733.	
PR	02-DEC-1999; 99WO-US28551.	
PR	22-DEC-1999; 99WO-US30720.	
PR	01-MAR-2000; 2000WO-US05601.	
PR	02-MAR-2000; 2000WO-US05841.	
PR	22-MAY-2000; 2000WO-US14042.	
PR	02-JUN-2000; 2000WO-US15264.	
PR	23-AUG-2000; 2000WO-US23522.	
PR	24-AUG-2000; 2000WO-US23228.	
PR	01-DEC-2000; 2000WO-US32678.	
PR	20-DEC-2000; 2000WO-US34956.	
PR	28-FEB-2001; 2001WO-US06520.	
PR	01-JUN-2001; 2001WO-US17800.	
PR	20-JUN-2001; 2001WO-US19692.	
PR	29-JUN-2001; 2001WO-US21066.	
PR	09-JUL-2001; 2001WO-US21735.	
PR	15-MAY-1998; 98US-085579P.	
PR	15-DEC-1998; 98US-112514P.	
PR	22-DEC-1998; 98US-113300P.	
PR	23-DEC-1998; 98US-113430P.	
PR	23-DEC-1998; 98US-113605P.	
PR	23-DEC-1998; 98US-113621P.	
PR	23-DEC-1998; 98US-114140P.	
PR	12-JAN-1999; 99US-115552P.	
PR	22-JAN-1999; 99US-116843P.	
PR	23-MAR-1999; 99US-125774P.	
PR	23-MAR-1999; 99US-125778P.	
PR	24-MAR-1999; 99US-125828P.	
PR	31-MAR-1999; 99US-127035P.	
PR	05-APR-1999; 99US-127706P.	
PR	13-APR-1999; 99US-129122P.	
PR	21-APR-1999; 99US-130359P.	
PR	27-APR-1999; 99US-131270P.	
PR	27-APR-1999; 99US-131272P.	
PR	27-APR-1999; 99US-131291P.	
PR	04-MAY-1999; 99US-132371P.	
PR	04-MAY-1999; 99US-132379P.	
PR	04-MAY-1999; 99US-132383P.	
PR	25-MAY-1999; 99US-135750P.	
PR	08-JUN-1999; 99US-138166P.	
PR	20-JUL-1999; 99US-144731P.	
PR	03-AUG-1999; 99US-146970P.	
PR	29-OCT-1999; 99US-162506P.	
PR	16-AUG-2001; 2001US-0931836.	
XX		
PA	(GETH) GENENTECH INC.	
XX		
XX		
PI	Deanoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J,	
PI	Stewart TA, Watanabe CK, Wood WI, Zhang Z;	
XX		
DR	WPI; 2003-341962/32.	
DR	N-PSDB; ACA06173.	
XX		
XX		
PT	Novel isolated PRO polypeptides e.g., PRO4334, PRO1122, PRO1889,	
PT	PRO1890, PRO1887, PRO1785, PRO4353, useful for treating sports	
PT	injuries, arthritis, diabetes, obesity, hyper- or hypo-insulinaemia	
XX		
XX	Claim 12; Fig 36; 194pp; English.	
XX		
CC	The present invention relates to the isolation of novel human PRO	
CC	polypeptides, and the polynucleotide sequences encoding them. The	
CC	PRO polypeptides are secreted and transmembrane proteins. The PRO	
CC	polypeptides and polynucleotides are useful in diagnosing or	
CC	treating various bone and/or cartilage disorders (e.g. sports	
CC	injuries, arthritis), various insulin deficient states (e.g. diabetes	

CC mellitus, hypo-insulinaemia), obesity, hyper-insulinaemia,
CC haemoglobin-associated disorders (e.g. thalassemias), kidney disorders
CC associated with decreased mesangial cell function (e.g. Berger disease),
CC or other nephropathies associated with Schönlein-Henoch purpura,
CC celiac disease, dermatitis herpetiformis or Crohn's disease. The PRO
CC polynucleotide sequences may be used as hybridisation probes in
CC chromosome and gene mapping, or in generating antisense RNA and DNA.
CC They are also useful in preparing PRO polypeptides, in assays to
CC identify other proteins or molecules involved in binding reaction, to
CC generate transgenic animals or knockout animals, which in turn are
CC useful in the development and screening of therapeutically useful
CC reagents, for chromosome identification, and tissue typing. The PRO
CC polypeptides and nucleic acid molecules are also useful in gene
CC therapy, and as molecular weight markers for protein electrophoresis
CC purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC PRO polypeptides, or for the affinity purification of the polypeptides
CC from recombinant cell culture or natural sources. ABU69103-ABU69125
CC represent the human PRO polypeptides of the invention.

XX SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGFLGTGFWILVLPLOAFPKPGSQDKSLHNRLSAERPLNQLIAEAEDKIKKTYPP 60
Db 1 MGFLGTGFWILVLPLOAFPKPGSQDKSLHNRLSAERPLNQLIAEAEDKIKKTYPP 60
Qy 61 ENKPGQSNYSFVDNLNLLKATEKEKTEKROSTRSPDLNKLNVEDVSTKPKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKATEKEKTEKROSTRSPDLNKLNVEDVSTKPKLIDDY 120
Qy 121 DSTKSGLDHKQDDPDGLHQLDGLPTLTAEDIVHKIAARIYEENDRAVDFDKIVSKLNLGL 180
Db 121 DSTKSGLDHKQDDPDGLHQLDGLPTLTAEDIVHKIAARIYEENDRAVDFDKIVSKLNLGL 180
Qy 181 ITESQAHTLEDEAVELQKLSKEANNYEEDPNKPTSWTENQAQKIPEKVTPMMAIQDGL 240
Db 181 ITESQAHTLEDEAVELQKLSKEANNYEEDPNKPTSWTENQAQKIPEKVTPMMAIQDGL 240
Qy 241 AKGENDETSTNTLTNTGLERTTYSSEDFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
Db 241 AKGENDETSTNTLTNTGLERTTYSSEDFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
Qy 301 TIMKTLIDFVQMWKYGTISPSEGVSVLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVQMWKYGTISPSEGVSVLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
Qy 361 KSHEETDSTKEEAARMEYKSLKSDTKDDNSNPGKTDPEKGTAYLEAIRKNIEWLK 420
Db 361 KSHEETDSTKEEAARMEYKSLKSDTKDDNSNPGKTDPEKGTAYLEAIRKNIEWLK 420
Qy 421 KHKKGKKNEDYDLKMRDFINKQADAYVEKGIIDKEEAIAIKRIYSSL 468
Db 421 KHKKGKKNEDYDLKMRDFINKQADAYVEKGIIDKEEAIAIKRIYSSL 468

RESULT 7

ABU71580
ID ABU71580 standard; Protein; 468 AA.
XX AC ABU71580;
XX AC
DT 10-JUN-2003 (first entry)
XX DE Human secreted polypeptide PR05990.
XX KW Human; gene therapy; tumour; cancer.
XX OS Homo sapiens.
XX PN US2003013855-A1.

XX 16-JAN-2003.
XX 03-MAY-2002; 2002US-0063616.
XX 30-DEC-1998; 98KB-0062142.
XX 08-MAR-1999; 99WO-US05028.
XX 14-MAY-1999; 99WO-US10733.
XX 30-DEC-1999; 99WO-US31274.
XX 18-FEB-2000; 2000WO-US04341.
XX 01-MAR-2000; 2000WO-US05601.
XX 02-MAR-2000; 2000WO-US05841.
XX 21-MAR-2000; 2000WO-US07532.
XX 22-MAY-2000; 2000WO-US14042.
XX 02-JUN-2000; 2000WO-US15264.
XX 24-AUG-2000; 2000WO-US23328.
XX 10-NOV-2000; 2000WO-US30873.
XX 01-DEC-2000; 2000WO-US32678.
XX 20-DEC-2000; 2000WO-US34956.
XX 28-FEB-2001; 2001WO-US06520.
XX 01-JUN-2001; 2001WO-US17800.
XX 14-MAY-1999; 99US-0311832.
XX 25-AUG-1999; 99US-0380137.
XX 25-AUG-1999; 99US-0380138.
XX 25-AUG-1999; 99US-0380139.
XX 25-AUG-1999; 99US-0380142.
XX 15-SEP-1999; 99US-0397342.
XX 18-OCT-1999; 99US-0403297.
XX 12-NOV-1999; 99US-0423844.
XX 22-AUG-2000; 2000US-0644848.
XX 18-SEP-2000; 2000US-0664610.
XX 18-SEP-2000; 2000US-0665350.
XX 08-NOV-2000; 2000US-0709238.
XX 20-DEC-2000; 2000US-0747259.
XX 22-MAR-2001; 2001US-0816744.
XX 10-MAY-2001; 2001US-0854208.
XX 10-MAY-2001; 2001US-0854280.
XX 30-MAY-2001; 2001US-0870574.
XX 05-JUN-2001; 2001US-0874503.
XX 29-JUN-2001; 2001US-0869599.
XX 18-JUL-2001; 2001US-0908827.
XX 06-DEC-2001; 2001US-0006867.

(GETH) GENENTECH INC.

Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;

WPI; 2003-330485/31.

N-PSDB; ACAS8884.

New isolated antibody specifically binding a PRO polypeptide, useful
for the preparation of a medicament for treating disorders with the
aberrant expression or activity of the PRO polypeptide, such as tumor
conditions and cancer

Example 17; Page 218-219; 406pp; English.

The invention relates to an antibody that binds to a polypeptide with a
fully defined sequence given in the specification. The methods and
compositions (containing antibodies that specifically bind a PRO
polypeptide) of the present invention are useful for the preparation of a
medicament for the treatment of disorders associated with the aberrant
expression or activity of the PRO polypeptide, such as tumour conditions
and cancer. They can also be used to generate transgenic or knockout
animals useful in the development and screening of therapeutically useful
reagents. The PRO polypeptides and encoding nucleic acids can be used as
molecular weight markers for protein electrophoresis, chromosome
identification and tissue typing. The PRO polypeptides are useful to
induce angiogenesis e.g wound healing; in the treatment of sports-related
joint problems, articular cartilage defects, osteoarthritis or rheumatoid
arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The
antibodies may be used in various diagnostic, competitive binding and/or

DE	Human PRO polypeptide #75.	
XX	Human; PRO polypeptide; secreted and transmembrane protein;	
KW	anti-PRO antibody; diagnostic assay; gene expression.	
XX	Homo sapiens.	
OS		
XX	US2003023042-A1.	
PN	30-JAN-2003.	
XX		
PD	01-MAY-2002; 2002US-0063502.	
XX		
PF	06-DEC-2001; 2001US-0006867.	
XX		
PR	(GETH) GENENTECH INC.	
XX	Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PU;	
PI	Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;	
XX	WPI; 2003-331484/31.	
DR	N-ESDB; ACA63447.	
XX		
XX	Novel monoclonal antibody that binds to secreted and transmembrane	
PT	polypeptide, useful for detecting and purifying the polypeptide and	
PT	also for treating conditions responsive to the antibody -	
XX	Disclosure; Fig 150; 408pp; English.	
PS		
XX		
CC	The present invention relates to the isolation of novel human PRO	
CC	polypeptides, and the polynucleotide sequences encoding them. The	
CC	PRO polypeptides are secreted and transmembrane proteins. The PRO	
CC	polypeptides and polynucleotides are useful for preparing a	
CC	medicament useful in the treatment of a condition responsive to	
CC	anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic	
CC	assays for PRO, by detecting its expression in specific cells,	
CC	tissues or serum, and for affinity purification of PRO from	
CC	recombinant cell culture or natural sources. ABU72109-ABU72192	
CC	represent the human PRO polypeptides of the invention.	
XX		
SQ	Sequence 468 AA;	
Query Match 100.0%; Score 2399; DB 24; Length 468;		
Best Local Similarity 100.0%; Pred. No. 4.1e-146;		
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 MGFLGTGTWILVLPIQAPPKPGSQDKSLHNRELSAERPLNEQIAEAEEDKIKTYPP	60
Db	1 MGFLGTGTWILVLPIQAPPKPGSQDKSLHNRELSAERPLNEQIAEAEEDKIKTYPP	60
Qy	61 ENKPGQSNYSFVDNLNLLKAITKEKIEKERQSISSPLDNKLNVEDVDSTKRNKLIDDY	120
Db	61 ENKPGQSNYSFVDNLNLLKAITKEKIEKERQSISSPLDNKLNVEDVDSTKRNKLIDDY	120
Qy	121 DSTKSGLDHKFPDQDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFDKIVSKLNLGL	180
Db	121 DSTKSGLDHKFPDQDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFDKIVSKLNLGL	180
Qy	181 ITESQAHTLEDEVAEVLQKLISKANNYEEDPNKPTSWTEQAQKIEPKVTPMAAIQDGL	240
Db	181 ITESQAHTLEDEVAEVLQKLISKANNYEEDPNKPTSWTEQAQKIEPKVTPMAAIQDGL	240
Qy	241 AKGENDETVSNTLTNTGLERTKTYSEDNPEELQYFPNFPYALLKSIDSEKAKEKETLI	300
Db	241 AKGENDETVSNTLTNTGLERTKTYSEDNPEELQYFPNFPYALLKSIDSEKAKEKETLI	300
Qy	301 TIMKTLIDFVQVMVKYGTIGPEGVSYLENLDEMIALQTNKLEKNAATDNISKLPAPSE	360
Db	301 TIMKTLIDFVQVMVKYGTIGPEGVSYLENLDEMIALQTNKLEKNAATDNISKLPAPSE	360
Qy	361 KSHETDSTKEAAKMEKEVGSLSKDTKDNNSPGGKTDEPKGKTEYLAIRKNIEWLK	420
Db	361 KSHETDSTKEAAKMEKEVGSLSKDTKDNNSPGGKTDEPKGKTEYLAIRKNIEWLK	420

Qy	421 KHDKKGKVEDYDLSKMRDFINKQADAYVEKGILDKEEAIAIKRIYSSL	468
Db	421 KHDKKGKVEDYDLSKMRDFINKQADAYVEKGILDKEEAIAIKRIYSSL	468
RESULT 10		
ABU66614		
ID	ABU66614 standard; Protein; 468 AA.	
XX		
AC	ABU66614;	
XX		
DT	23-MAY-2003 (first entry)	
XX		
DE	Human PRO polypeptide #45.	
XX		
KW	Human; PRO polypeptide; secreted and transmembrane protein;	
KW	tumour necrosis factor-alpha; TNF-alpha; blood; proliferation;	
KW	differentiation; chondrocyte; tumour; genetic disorder;	
KW	cytostatic.	
XX		
OS	Homo sapiens.	
XX		
PN	US2003036180-A1.	
XX		
PD	20-FEB-2003.	
XX		
PF	09-MAY-2002; 2002US-0143114.	
XX		
PR	31-MAR-1997; 97WO-US05230.	
PR	12-JUN-1998; 98WO-US12456.	
PR	14-JUL-1998; 98WO-US14552.	
PR	28-AUG-1998; 98WO-US17888.	
PR	10-SEP-1998; 98WO-US18824.	
PR	14-SEP-1998; 98WO-US19093.	
PR	14-SEP-1998; 98WO-US19094.	
PR	14-SEP-1998; 98WO-US19177.	
PR	16-SEP-1998; 98WO-US19330.	
PR	17-SEP-1998; 98WO-US19437.	
PR	07-OCT-1998; 98WO-US21141.	
PR	29-OCT-1998; 98WO-US22991.	
PR	23-OCT-1998; 98WO-US22992.	
PR	20-NOV-1998; 98WO-US24855.	
PR	01-DEC-1998; 98WO-US25108.	
PR	05-JAN-1999; 99WO-US00106.	
PR	08-MAR-1999; 99WO-US05028.	
PR	10-MAR-1999; 99WO-US05190.	
PR	20-APR-1999; 99WO-US08615.	
PR	14-MAY-1999; 99WO-US10733.	
PR	02-JUN-1999; 99WO-US12252.	
PR	01-SEP-1999; 99WO-US20111.	
PR	08-SEP-1999; 99WO-US20594.	
PR	13-SEP-1999; 99WO-US20944.	
PR	15-SEP-1999; 99WO-US21090.	
PR	05-OCT-1999; 99WO-US21547.	
PR	29-NOV-1999; 99WO-US28214.	
PR	30-NOV-1999; 99WO-US28313.	
PR	30-NOV-1999; 99WO-US28409.	
PR	01-DEC-1999; 99WO-US28301.	
PR	01-DEC-1999; 99WO-US28634.	
PR	02-DEC-1999; 99WO-US28551.	
PR	02-DEC-1999; 99WO-US28565.	
PR	16-DEC-1999; 99WO-US30095.	
PR	20-DEC-1999; 99WO-US30911.	
PR	22-DEC-1999; 99WO-US30999.	
PR	30-DEC-1999; 99WO-US30720.	
PR	30-DEC-1999; 99WO-US31243.	
PR	03-JAN-2000; 99WO-US31274.	
PR	06-JAN-2000; 2000WO-US00219.	
PR	06-JAN-2000; 2000WO-US00277.	
PR	06-JAN-2000; 2000WO-US00376.	

XX OS Homo sapiens.
XX PN US2003032155-A1.
XX PD 13-FEB-2003.
XX PF 03-MAY-2002; 2002US-0137865.
XX PR 31-MAR-1997; 97WO-US05230.
XX PR 12-JUN-1998; 98WO-US12456.
XX PR 14-JUL-1998; 98WO-US14552.
XX PR 28-AUG-1998; 98WO-US17888.
XX PR 10-SEP-1998; 98WO-US18824.
XX PR 14-SEP-1998; 98WO-US19093.
XX PR 14-SEP-1998; 98WO-US19177.
XX PR 16-SEP-1998; 98WO-US19330.
XX PR 17-SEP-1998; 98WO-US19437.
XX PR 07-OCT-1998; 98WO-US21141.
XX PR 29-OCT-1998; 98WO-US22991.
XX PR 29-OCT-1998; 98WO-US22992.
XX PR 20-NOV-1998; 98WO-US24855.
XX PR 01-DEC-1998; 98WO-US25108.
XX PR 05-JAN-1999; 99WO-US00106.
XX PR 08-MAR-1999; 99WO-US05028.
XX PR 10-MAR-1999; 99WO-US05190.
XX PR 20-APR-1999; 99WO-US08615.
XX PR 14-MAY-1999; 99WO-US10733.
XX PR 02-JUN-1999; 99WO-US12252.
XX PR 01-SEP-1999; 99WO-US20111.
XX PR 08-SEP-1999; 99WO-US20594.
XX PR 13-SEP-1999; 99WO-US20944.
XX PR 15-SEP-1999; 99WO-US21090.
XX PR 15-SEP-1999; 99WO-US21547.
XX PR 05-OCT-1999; 99WO-US23089.
XX PR 29-NOV-1999; 99WO-US28214.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 30-NOV-1999; 99WO-US28409.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 01-DEC-1999; 99WO-US28634.
XX PR 02-DEC-1999; 99WO-US28551.
XX PR 02-DEC-1999; 99WO-US28564.
XX PR 02-DEC-1999; 99WO-US28565.
XX PR 16-DEC-1999; 99WO-US30095.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 20-DEC-1999; 99WO-US30999.
XX PR 22-DEC-1999; 99WO-US30720.
XX PR 30-DEC-1999; 99WO-US31243.
XX PR 30-DEC-1999; 99WO-US31274.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PR 06-JAN-2000; 2000WO-US00277.
XX PR 06-JAN-2000; 2000WO-US00376.
XX PR 11-FEB-2000; 2000WO-US03565.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 18-FEB-2000; 2000WO-US04342.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 24-FEB-2000; 2000WO-US04914.
XX PR 24-FEB-2000; 2000WO-US05004.
XX PR 01-MAR-2000; 2000WO-US05601.
XX PR 02-MAR-2000; 2000WO-US05746.
XX PR 10-MAR-2000; 2000WO-US05841.
XX PR 10-MAR-2000; 2000WO-US06319.
XX PR 15-MAR-2000; 2000WO-US06884.
XX PR 20-MAR-2000; 2000WO-US07377.
XX PR 21-MAR-2000; 2000WO-US07532.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 17-MAY-2000; 2000WO-US13705.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 30-MAY-2000; 2000WO-US14941.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 28-JUL-2000; 2000WO-US20710.
XX PR 11-AUG-2000; 2000WO-US22031.
XX PR 23-AUG-2000; 2000WO-US23522.
XX PR 24-AUG-2000; 2000WO-US23328.
XX PR 08-NOV-2000; 2000WO-US30952.
XX PR 10-NOV-2000; 2000WO-US30873.
XX PR 01-DEC-2000; 2000WO-US32678.
XX PR 20-DEC-2000; 2000WO-US34956.
XX PR 28-FEB-2001; 2001WO-US06520.
XX PR 01-MAR-2001; 2001WO-US06666.
XX PR 25-MAY-2001; 2001WO-US17092.
XX PR 01-JUN-2001; 2001WO-US17800.
XX PR 20-JUN-2001; 2001WO-US19692.
XX PR 22-JUN-2001; 2001WO-US20116.
XX PR 29-JUN-2001; 2001WO-US21086.
XX PR 09-JUL-2001; 2001WO-US21735.
XX PR 20-DEC-2000; 2000US-0747259.
XX PR 28-FEB-2001; 2001US-0796498.
XX PR 09-MAR-2001; 2001US-0802706.
XX PR 14-MAR-2001; 2001US-0808689.
XX PR 22-MAR-2001; 2001US-0816744.
XX PR 05-APR-2001; 2001US-0828366.
XX PR 10-MAY-2001; 2001US-0854208.
XX PR 10-MAY-2001; 2001US-0854280.
XX PR 18-MAY-2001; 2001US-0860216.
XX PR 25-MAY-2001; 2001US-0866028.
XX PR 25-MAY-2001; 2001US-0866034.
XX PR 01-JUN-2001; 2001US-0872035.
XX PR 05-JUN-2001; 2001US-0874503.
XX PR 14-JUN-2001; 2001US-0882636.
XX PR 19-JUN-2001; 2001US-0886342.
XX PR 21-JUN-2001; 2001US-0887879.
XX PR 18-JUL-2001; 2001US-0908827.
XX PR 06-AUG-2001; 2001US-0924419.
XX PR 09-AUG-2001; 2001US-0927796.
XX PR 16-AUG-2001; 2001US-0931836.
XX PR 19-DEC-2001; 2001US-0028072.
XX PA (GETH) GENENTECH INC.
XX XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
XX PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
XX PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX XX WPI; 2003-331925/31.
XX DR N-PSDB; ACA04068.
XX XX New secreted and transmembrane nucleic acids and polypeptides.
XX PT designated as PRO, useful for treating inflammation, organ failure,
XX PT atherosclerosis, cardiac injury, infertility, birth defects, premature
XX PT aging, AIDS, or cancer
XX XX Claim 12; Fig 90; 659pp; English.
XX PS The invention relates to an isolated nucleic acid comprising, or which is
XX CC at least 80% identical to, or the full-length coding sequence of, any of
XX CC the 275 nucleotide sequences, encoding the corresponding PRO polypeptide
XX CC (one of 275 secreted or transmembrane proteins). The nucleic acid
XX CC further comprises the full-length coding sequence of the DNA deposited
XX CC under American Type Culture Collection (ATCC) accession number in a list
XX CC given in the specification. Also included are vectors and host
XX CC cells for producing PRO proteins, PRO fusion proteins, anti-PRO
XX CC antibodies, PRO extracellular domains and mature sequences, methods
XX CC of detecting PRO proteins, methods for stimulating the release of
XX CC TNF-alpha (tumour necrosis factor alpha) from human blood,
XX CC (and the proliferation of differentiation of chondrocyte cells, the
XX CC proliferation of, or gene expression in pericyte cells, the release or
XX CC proteoglycans from cartilage, proliferation of inner ear utricular
XX CC supporting cells, the proliferation of T-lymphocyte cells, the release
XX CC of a cytokine from peripheral blood mononuclear cells (PBMC), or the
XX CC proliferation of endothelial cells), a method for modulating the uptake
XX CC of glucose or free fatty acid (FFA) by skeletal muscle cells,
XX CC a method for inhibiting the binding of A-peptide to factor VIIA,
XX CC or the differentiation of adipocyte cells, a method for detecting the
XX CC presence of a tumour in a mammal and an oligonucleotide probe derived

CC from any of the nucleotide sequences cited above. The nucleic acids and
 CC polypeptides are useful for treating inflammatory diseases, organ
 CC failure, atherosclerosis, cardiac injury, infertility, birth defects,
 CC premature aging, AIDS (acquired immunodeficiency syndrome), cancer, or
 CC diabetic complications. The nucleic acids are useful as hybridisation
 CC probes, in chromosome and gene mapping, and in generating antisense RNA
 CC or DNA. The polypeptides are useful as pharmaceuticals, diagnostics,
 CC biosensors or bioreactors. Both are useful in tissue typing.
 CC The present sequence represents a PRO protein of the invention.
 XX
 SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
 Best Local Similarity 100.0%; Pred. No. 4.1e-146;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MGFLGTGWLVLPIQAPKPGSGDSKSLHRELGAERPLNEQIAEAEDKIKTYPP 60
 Db 1 MGFLGTGWLVLPIQAPKPGSGDSKSLHRELGAERPLNEQIAEAEDKIKTYPP 60
 Qy 61 ENKPGQSNYSFVNLNLKATTEKEKIEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
 Db 61 ENKPGQSNYSFVNLNLKATTEKEKIEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
 Qy 121 DSTKSGLDHFKPDGPHQDGLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
 Db 121 DSTKSGLDHFKPDGPHQDGLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
 Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKPEKVTPTMAAQDGL 240
 Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKPEKVTPTMAAQDGL 240
 Qy 241 AKGENDETQSVNTLTNGLERRTKTYSSEDFEELQYFPNFVALLKSIDSEKEAKEKETLI 300
 Db 241 AKGENDETQSVNTLTNGLERRTKTYSSEDFEELQYFPNFVALLKSIDSEKEAKEKETLI 300
 Qy 301 TIMKTLIDFVKMVKYGTISPFGVSVLENLDEMIALQTNKLEKATDNISKLPAPSE 360
 Db 301 TIMKTLIDFVKMVKYGTISPFGVSVLENLDEMIALQTNKLEKATDNISKLPAPSE 360
 Qy 361 KSHEETDSTKEEAARKEKEYGSLKSDTNDSPGKTDPEKGTAYLEAIRKNIEWLK 420
 Db 361 KSHEETDSTKEEAARKEKEYGSLKSDTNDSPGKTDPEKGTAYLEAIRKNIEWLK 420
 Qy 421 KHKKGKNEDYDLSKMRDFINKQADAYVEKGILDKGEAEAKRIYSSL 468
 Db 421 KHKKGKNEDYDLSKMRDFINKQADAYVEKGILDKGEAEAKRIYSSL 468

RESULT 12
 ABUS9695
 ID ABUS9695 standard; Protein; 468 AA.
 XX
 AC ABUS9695;
 XX
 DT 13-MAY-2003 (first entry)
 XX
 DE Novel secreted and transmembrane protein PRO5990.
 XX
 KW Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;
 KW cardiac insufficiency disorder; cancer; tumour; immune response;
 KW adrenal cortical capillary endothelial growth; c-fos induction;
 KW vascular endothelial growth factor inhibition; VEGF inhibition;
 KW endothelial cell growth inhibitor; T-lymphocytes stimulation;
 KW retinal neurons cell survival; rod photoreceptor cell survival;
 KW retinal disorder; retinitis pigmentosa; kidney disorder;
 KW mammalian kidney mesangial cell proliferation; Berger disease;
 KW dermatitis; herpeticiformis; Crohn's disease; chondrocyte proliferation;
 KW chondrocyte redifferentiation; sports injury; arthritis.
 XX
 OS Homo sapiens.
 XX
 PN US2003017563-A1.

XX
 PD 23-JAN-2003.
 XX
 XX 07-MAY-2002; 2002US-0140808.
 XX
 PR 31-MAR-1997; 97WO-US05230.
 PR 12-JUN-1998; 98WO-US12456.
 PR 14-JUL-1998; 98WO-US14552.
 PR 28-AUG-1998; 98WO-US17888.
 PR 10-SEP-1998; 98WO-US18824.
 PR 14-SEP-1998; 98WO-US19093.
 PR 14-SEP-1998; 98WO-US19094.
 PR 14-SEP-1998; 98WO-US19177.
 PR 16-SEP-1998; 98WO-US19330.
 PR 17-SEP-1998; 98WO-US19437.
 PR 07-OCT-1998; 98WO-US21141.
 PR 29-OCT-1998; 98WO-US22991.
 PR 29-OCT-1998; 98WO-US22992.
 PR 20-NOV-1998; 98WO-US24855.
 PR 01-DEC-1998; 98WO-US25108.
 PR 05-JAN-1999; 99WO-US00106.
 PR 08-MAR-1999; 99WO-US05028.
 PR 10-MAR-1999; 99WO-US05190.
 PR 20-APR-1999; 99WO-US08615.
 PR 14-MAY-1999; 99WO-US10733.
 PR 01-SEP-1999; 99WO-US12252.
 PR 01-SEP-1999; 99WO-US20111.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 30-NOV-1999; 99WO-US28409.
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 16-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 22-DEC-1999; 99WO-US30720.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05746.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 10-MAR-2000; 2000WO-US06319.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 11-AUG-2000; 2000WO-US22031.
 PR 23-AUG-2000; 2000WO-US23522.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 10-NOV-2000; 2000WO-US30873.

Db 220 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 279
Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTMPMAAIQDGL 240
Db 280 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTMPMAAIQDGL 339
Qy 241 AKGENDETVSNTLTNTGLERTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
Db 340 AKGENDETVSNTLTNTGLERTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 399
Qy 301 TIMKTLIDFVKMMVKYGTISPPEGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 400 TIMKTLIDFVKMMVKYGTISPPEGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 459
Qy 361 KSHEETDSTKEEAAKMEKEYGSLKDDSTKDDNSNPGKTDPEPKGTEAYLEAIRKNIEWLK 420
Db 460 KSHEETDSTKEEAAKMEKEYGSLKDDSTKDDNSNPGKTDPEPKGTEAYLEAIRKNIEWLK 519
Qy 421 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
Db 520 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 567

RESULT 15

AAW93910 ID AAW93910 standard; Protein; 468 AA.

XX AC AAW93910;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide, SEQ ID NO: 4061.

XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.

XX OS Homo sapiens.

XX PN EP1130094-A2.

XX PD 05-SEP-2001.

XX PF 07-JUL-2000; 2000EP-0114089.

XX PR 08-JUL-1999; 99JP-0194486.

XX PR 11-JAN-2000; 2000JP-0118774.

XX PR 02-MAY-2000; 2000JP-0183765.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX DR WPI; 2001-524255/58.

XX DR N-PSDB; AAK94872.

XX PT 830 Primers useful for synthesizing full length cDNA clones and their

XX PT use in genetic manipulation -

XX PS Claim 8; SEQ ID NO 4061; 1380pp + sequence listing; English.

XX CC The invention relates to primers for synthesising full length cDNA

XX CC clones. 830 cDNA molecules encoding a human protein have been

XX CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

XX CC molecules have been determined. Primers for synthesising the full length

XX CC cDNA are useful for clarifying the function of the protein encoded by

XX CC the cDNA. The full length clones were obtained by construction of full

XX CC length enriched cDNA libraries that were synthesised by the oligo-capping

XX CC method. The primers enable the production of the full length cDNA easily

XX CC without any special methods. The present sequence is a polypeptide

XX CC encoded by a full length human cDNA of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in CD-ROM format directly from EPO.

XX SQ Sequence 468 AA;
Query Match 99.7%; Score 2392; DB 22; Length 468;
Best Local Similarity 99.8%; Pred. No. 1.2e-145;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MGFLGTGTWILVLVLPQAFPKPGSQDKSLHNRELSAERPLNEQIAEABEDKIKKTYPP 60
Db 1 MGFLGTGTWILVLVLPQAFPKPGSQDKSLHNRELSAERPLNEQIAEABEDKIKKTYPP 60
Qy 61 ENKPGQSNYSFVONLNLKKAITEKEKIEKERQSIKSSPLDNKLNVEDVDSTKNKRLIDDY 120
Db 61 ENKPGQSNYSFVONLNLKKAITEKEKIEKERQSIKSSPLDNKLNVEDVDSTKNKRLIDDY 120
Qy 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTMPMAAIQDGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTMPMAAIQDGL 240
Qy 241 AKGENDETVSNTLTNTGLERTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
Db 241 AKGENDETVSNTLTNTGLERTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
Qy 301 TIMKTLIDFVKMMVKYGTISPPEGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKMMVKYGTISPPEGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360
Qy 361 KSHEETDSTKEEAAKMEKEYGSLKDDSTKDDNSNPGKTDPEPKGTEAYLEAIRKNIEWLK 420
Db 361 KSHEETDSTKEEAAKMEKEYGSLKDDSTKDDNSNPGKTDPEPKGTEAYLEAIRKNIEWLK 420
Qy 421 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
Db 421 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468

Search completed: January 2, 2004, 12:34:22

Job time : 47 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 21 Seconds

(without alignments)
942.928 Million cell updates/sec

Title: ABU66614

Perfect score: 2399

Sequence: 1 MGFLGTGWLVLVLPQAF.....EKGILDKBAEAKRIYSSL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*

5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pap.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179	7.5	1786	3	US-08-973-462-8
2	172.5	7.2	1164	3	US-08-923-992A-2
3	169	7.0	1098	3	US-08-923-992A-8
4	168	7.0	1164	3	US-08-923-992A-10
5	167.5	7.0	957	4	US-09-914-259-16
6	166	6.9	1104	3	US-08-923-992A-4
7	166	6.9	1128	3	US-08-923-992A-6
8	166	6.9	2662	4	US-09-595-684B-31
9	163.5	6.8	956	4	US-09-914-259-17
10	162	6.8	8991	4	US-08-714-741-32
11	159.5	6.6	816	2	US-08-533-306A-6
12	159.5	6.6	816	2	US-08-742-923A-6
13	158.5	6.6	885	2	US-08-533-306A-4
14	158.5	6.6	885	2	US-08-742-923A-4
15	156.5	6.5	1087	4	US-09-914-259-12
16	156	6.5	1354	3	US-08-685-871-2
17	154.5	6.4	1231	4	US-08-714-741-41
18	153.5	6.4	984	1	US-08-242-932-2
19	153.5	6.4	984	1	US-08-714-481-2
20	153.5	6.4	984	5	PT-US95-06111-2
21	153	6.4	1588	5	PT-US93-07261-11
22	153	6.4	1663	5	PT-US93-07261-16
23	153	6.4	2285	4	US-09-308-375-2
24	150.5	6.3	1010	4	US-09-134-001C-5178
25	147	6.1	534	4	US-09-103-664A-2
26	146.5	6.1	1388	2	US-08-685-576-1
27	146.5	6.1	1388	2	US-08-685-576-4

28	145.5	6.1	652	4	US-08-559-896B-2	Sequence 2, Appli
29	145	6.0	588	4	US-08-714-741-42	Sequence 42, Appl
30	144.5	6.0	343	3	US-08-937-271-17	Sequence 17, Appl
31	144	6.0	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
32	143.5	6.0	864	4	US-08-714-741-40	Sequence 40, Appl
33	143	6.0	3878	4	US-09-914-259-11	Sequence 11, Appl
34	142	5.9	967	4	US-09-914-259-21	Sequence 21, Appl
35	142	5.9	1032	4	US-09-914-259-26	Sequence 26, Appl
36	141.5	5.9	1146	4	US-08-914-999-6	Sequence 6, Appli
37	141	5.9	1507	3	US-08-929-329-5	Sequence 5, Appli
38	139.5	5.8	963	4	US-09-914-259-20	Sequence 20, Appl
39	139	5.8	1939	4	US-09-310-187A-1	Sequence 1, Appli
40	139	5.8	3111	2	US-08-460-309-4	Sequence 4, Appli
41	139	5.8	3111	2	US-08-125-077-4	Sequence 4, Appli
42	138	5.8	630	3	US-08-973-462-9	Sequence 9, Appli
43	138	5.8	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
44	138	5.8	1234	4	US-09-592-054-8	Sequence 8, Appli
45	137.5	5.7	586	2	US-08-630-822A-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1

US-08-973-462-8

; Sequence 8, Application US/08973462B

; Patent No. 6191270

; GENERAL INFORMATION:

; APPLICANT: DRUILHE, PIERRE

; APPLICANT: DAUBERSIES, PIERRE

; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

; FILE REFERENCE: 0660-0125-0 PCT

; CURRENT APPLICATION NUMBER: US/08/973.462B

; CURRENT FILING DATE: 1998-02-06

; EARLIER APPLICATION NUMBER: PCT/FR96/00894

; EARLIER FILING DATE: 1996-06-12

; EARLIER APPLICATION NUMBER: FR 95/07007

; EARLIER FILING DATE: 1995-06-13

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 1786

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide

US-08-973-462-8

Query Match

Best Local Similarity 7.5%; Score 179; DB 3; Length 1786;

Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;

Qy 27 QDKSLHNRISAE--RPLNEQIAEAEEDKIKTYPPENKPGQSN-----YSFVDNL-- 75

Db 957 EENAVESNENVAENLEKLNFTVNTLDKVEETVEISGESLENNEMDKAFSEIFDNVKG 1016

Qy 76 ---NLKK-----AITEKEKIEKROSIRSSPLDNKLN-----VEDVDSTK-- 112

Db 1017 IQENLLTGMFRSTIETSIQSEKVDL--NENVSSILDNIENKKEGLLNKLENISSTEGV 1075

Qy 113 -----NKKLIDDDY-----STKGGLDHKFODDPDGLHOLGDTPLTAEI- 151

Db 1076 QETVTEHVEQNVYVDVDPAMKQDFLGILNEAGLKEMFNFLED--VPKSESDVITVEIK 1134

Qy 152 -----VHKIARIYEENDRAVDFKIVKLLNGLITESQAHTEDEVAEVLQKLSKE 204

Db 1135 DEPVQKEVEKETVSIIEEMEENIVD-----VLEBEKDLTKMDAVEESIEIS 1183

Qy 205 ANNYEDPNKPTSTMTENOAGKIPKVTPTMAAIQDGLAKGENDETSTVNTLTITNGLERRTK 264

Db 1184 SDSKEE-----TES-----IKDKKDVSLVVEEVDNDDESVEKVLKKNBEELMK 1231

Qy 265 TYSEDN-----FEELOYFFNFYA--LLKSIDSEKAK-----EKETLITIMK 304

Db 1232 DAVEINDITSKLIETQELNEVEADLIKOMEKLEKALSEDSKEIIDAKDDTLEKVIK 1291
 QY 305 -----TLIDFVKMVKYGTISPE--EGSVYLENDDEMIALQTK--NKLEKNATDNISKL 354
 Db 1292 EHDITTTILDEV---VELKQVEEDKIEKVSOLKLEEDILKEVKEIKELSEILEYKEL 1348
 QY 355 PPAPSEKSHETDSTKEEAAKMEKEYGSLKSTKDDNSNPGKTDPEPKGTEAYLEAIRK 414
 Db 1349 KTIETDILLEKEIEKHDFEFEEAEIKOLEADILKEVSSLEVEBEKKLEEVHE--LKE 1407
 QY 415 NIEWLKKHDK--KGNKEDYDLSKMRDFINKQADAY---VEKGILDK-----EEA 458
 Db 1408 EVEHIISDAHIKGLIED--DLEEVDDLKSGILDMKGMELGMDMKESLEDVTTTKLGERV 1466
 QY 459 EAIKRIYSS 467
 Db 1467 ESKDVLSS 1475

RESULT 2
 US-08-923-992A-2
 ; Sequence 2, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IGA Fc Binding forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,992A
 ; FILING DATE: 05-SEP-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/024,707
 ; FILING DATE: 06-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 ; TELEPHONE: (202) 371-2540
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1164 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-923-992A-2
 ;
 ; Query Match 7.2%; Score 172.5; DB 3; Length 1164;
 ; Best Local Similarity 21.6%; Pred. No. 0.0002;
 ; Matches 124; Conservative 77; Mismatches 189; Indels 185; Gaps 25;
 ;
 ; 36 LSARFLNE-----QTAAEDKIKTYPPENKPGQSNYSFVDNLLKAIT-EKEK 86
 ; 149 LELENQFNETNRLHLHIKHEEVEKDKAK-----QKTLKQSDTKVDLSNIDKLNHQKQSQ 204
 ; 87 IEK--ERQSTIRSSPLDNKL-----NVEDVSTKRNKLIDDYDSTKSLDHK 130

Db 205 VERMAQKGTINDEKDSMLKKIETDIRKQQAQDKKEDAEVKKVREELGKLFSSSTKAGLDQE 264
 QY 131 FQDDPDLHQLDGTPTABEDIVHKI-----AARIYENDRAVDFDKIVSKLLNLG 179
 Db 265 IQE-----HVKKET--SSBENTQKVDEHYANSIQNLAKSLEELDKATTNEQATQVKNQF 317
 QY 180 L-----ITESAHTLEDEVAEV-----LQKLSKE---ANNYEE 210
 Db 318 LENAQKLEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEG 377
 QY 211 DPNKPTSWTE-----NOAGKIPEKVTWMAAIOQDGLAKGENDETVPN 251
 Db 378 KLNQSKNLPELQLEBEAHSKLVQVEDFRKFKTSQVTPKKRVKRDLAANENNO---Q 434
 QY 252 TLTLTNGLERRTKTYSSEDFELQVFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVK 311
 Db 435 KIETVSPENITVYEGED-----VKFTVTAKS-DS-----KTTLDPSD 471
 QY 312 MMVKYG-TISPEGVSYLENLD-----EMIALQTK-----NKLEKNATD 349
 Db 472 LUTKYNPSVSDRISTNYKTNTDNHKAIEITIKNLKLNESQVTLKAKDDSGNVVEKTFPI 531
 QY 350 NISKLPAPSEKSHETDSTKEEAA-----KMEKEYGSLKD 385
 Db 532 TVQKKEEKQVKTPEQKDSKTEEKVPQBPQSKNDKNQQLQELIKSAQOELEKLEKAIKELME 591
 QY 386 STKDONSNGGKTDPEKGTAYLEAIRKNIEWLKK-----HDKGNKEDYDLSKMRDFIN 441
 Db 592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYKYTEHYFNKYKSDPMN 648
 QY 442 KOADAVVEKGILDKKEAE-----AIKRIYSS 467
 Db 649 YOLHAQME--MLTRKVVQYMKNPDAEIKKIFES 681
 ;
 ; RESULT 3
 ; US-08-923-992A-8
 ; Sequence 8, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IGA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,992A
 ; FILING DATE: 05-SEP-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/024,707
 ; FILING DATE: 06-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 ; TELEPHONE: (202) 371-2540
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 8:


```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-8

Query Match
Best Local Similarity 19.3%; Score 169; DB 3; Length 1098;
Matches 104; Conservative 80; Mismatches 182; Indels 174; Gaps 19;

Qy 25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSF---VDNLNLKAI 81
Db 74 GKREKQLQWKNLKNVDNHTLSHEQKNEFKTKIDETNDSALLELENFNETNRLHI 133
Qy 82 TEKEKIEKERSIRSSPL---DNKLNVEDVDSTKNR-----KLID-- 119
Db 134 KOHEVEKDKKAKQOKTLKQSDTKVDLSNIDKELNHQKSQAGITNEKDMSMLKIEDIR 193
Qy 120 -----YDSTKSGLDHFKQDDPDGGLHQLDGTPLTAEDIVHKI- 155
Db 194 KQAQOPDKKDAEVKVRBELGKLFSTKAGLDQEIQ-----HVKET--SSEENTQKVD 246
Qy 156 -----AARIYENDRAVDFKIVSKLNLGL-----ITSOA 186
Db 247 EHYANSLQNLAKSLEELDQATTNEQATQVKNQFLENAQKLEIQPLIKETNVKLYKAMS 306
Qy 187 HTLEDEVAEV-----LOKLISKE---ANNYEEDPNKPTSWTE----- 220
Db 307 ESLEQVEKELKHNSFANLELDIVAKSKEIVRYEGKLNQSKNLPKQLKEEAHKLQKV 366
Qy 221 ---NQAGKIPKVTMAAIOPLAGENDETVSNLTITLNGLERRTKTSYSDNFEELOYF 277
Db 367 EDPRKFKTSEQVTPKPKRLKRDLAANNQ---QKIETVSPENITVVEGED----- 415
Qy 278 PNFVALLKSISEKAKEKFTLITMTKLIDPFVQVMVYK--TISPEEGVSVLENLD-- 332
Db 416 VKFTVTAKS-DS-----KTLTDFSDLLTKYNFSDRISTNTKNTDNHKI 460
Qy 333 -----EMIALQTK---NKLEKNATDNISKLPAPSEKSHSEEDSTKEAAK 375
Db 461 AEITIKNLKNESQVTLKAKDGSNVVEKFTTITVQKKEQKVPKTEQKDSKTEKVP 520
Qy 376 MEKEYGSLKSTKDDNSPGKTDPEKGTBAYLEAIRKNIEWLKHDKKGNKEDYDLSK 435
Db 521 QEPK-----SNDKNQLOELIKSAQOQLEKLEKAIKELMEQPEIFSNPEYGTQK 568

RESULT 4
US-08-923-992A-10
; Sequence 10, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
```

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-10

Query Match
Best Local Similarity 19.0%; Score 168; DB 3; Length 1164;
Matches 121; Conservative 85; Mismatches 177; Indels 254; Gaps 25;

Qy 25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSF---VDNLNLKAI 81
Db 105 GKREKQLQWKNLKNVDNHTLSHEQKNEFKTKIDETNDSALLELENFNETNRLHI 164
Qy 82 TEKEKIEKERSIRSSPL---DNKLNVEDVDSTKNR-----K 115
Db 165 KOHEVEKDKKAKQOKTLKQSDTKVDLSNIDKELNHQKSQVETMAEQGITNEKDMSMLK 224
Qy 116 LID-----YDSTKSGLDHFKQDDPDGGLHQLDGTPLTAED 150
Db 225 KIEDIRKQAQADKEDAEVAVREELGKLFSTKAGLDQEIQ-----HVKET--TSEE 277
Qy 151 IVHKI-----AARIYENDRAVDFKIVSKLNL-----LGLTE-- 183
Db 278 NTQKVDHYHPNSLQNLAKSLEELDQATTNEQATQVKNQFLENAQKLEIQPLIKETNVK 337
Qy 184 -----SOAHTLEDEVAEVQKLISKE---ANNYEEDPNKPTSWTE-- 220
Db 338 LYKAMESLEQVEKQLKHNSQAN--LEDLVAK-----SKEIVRYEGKLNQSKNLPKQL 390
Qy 221 -----NQAGKIPKVTMAAIOPLAGENDE-----TVS--NTLTITNG 258
Db 391 LEEBAHKLQVVEDPRKFKTSEQVTPKRVKRDLAANNQKIELTVSPENITVVEG 450
Qy 259 LERTKITYSDNFELOYFPNFVALLKSISEKAKEKFTLITMTKLIDPFVQVMVYK- 317
Db 451 -----EDLK-----FTLTAKSDS-----KTLTDFSDLLTKYNP 478
Qy 318 TISPEEGVSVLENLD-----EMIALQTK---NKLEKNATDNISKLP 356
Db 479 SVSDRISTNTKNTDNHKAIEITIKNLKNESQVTLKAKDGSNVVQKFTTITVQKKEE 538
Qy 357 APSEKSHSEEDSTKEAAKMEKEYGSLKSTKDDNSPGKTDPEKGTBAYLEAIRKNI 416
Db 539 KOVPKTEQKDSKTEKVPQEPK-----SNDKNQLOELIKSAQOQLEKLEKAI 586
Qy 417 EWLKHKHDKGNKEDYDLSK-----MRDF 439
Db 587 KELMEQPEIFSNPEYGTQKSIWESQKEPIQEAITSFKKIIGDSSSKYVTEHYFNKYKSD 646
Qy 440 INKQADAYVEKGILDKEEA-----AIKRIYSS 467
Db 647 MNYQLHAQME--MLTRKVQVQYINKYPADNAEIKKIFES 681

RESULT 5
US-09-914-259-16
; Sequence 16, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
```

```

; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-16

Query Match          7.0%; Score 167.5; DB 4; Length 957;
Best Local Similarity 20.3%; Pred. No. 0.00036;
Matches 103; Conservative 98; Mismatches 182; Indels 125; Gaps 23;

QY 25 GSQDKSLNR-----ELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSFVDNLNL- 78
Db 321 QORAKTIKNTSVNLELTAE-----EMKKYKEKEKNTLKNVIOHLEMEL 367
QY 79 -----KATKEKIEKERSIRSPLDNKLNVEDVD-----STNRKLIIDYDSTKSG 126
Db 368 NRWRNGEAVPEDEQISAKDQK-NLEPCDNTPIIDNIAPVVGAGISTEEK--EKYDEISS 423
QY 127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAAIVEEND-----RAVFDKIVSKLLNLGLI 181
Db 424 LYRQLDDKDDINO-----QSQAELKIQQLDQELLASTRDYKEIQEELTRQIE 476
QY 182 TESQAHTLEDEVAEVLQKLSKEANNYE---EDPNKPTSWTNOAGKPIPEKVPMAAIQD 238
Db 477 NEA-----AKDEVEVLOAL-EELAVNDQKSQVEDKTRANEQLTDLAQKTTITTTQ 531
QY 239 GLAK-----GENDETVSNTLITNGLERR-----TK 264
Db 532 ELSQQLSLSNHQKRATEINLLKDLGEIGGIGTND---VKTLDVNGVIEEFETMAR 588
QY 265 TYSEDNFEELQYFNFYALLKS--IDSEKAKEKETLITIMTKLIDFVKMVKYGVITPSE 322
Db 589 LYISKMKSEVKSLSVNRSKQLESQMSDNKRNKASERELACQLLSIQHEAKISLT----- 644
QY 323 EGVSYLENDEMIALQYFNKLEK-----NATDNIKSLFPAPSEKSHETDSTKE----- 371
Db 645 ---DYMQNM-----QKRRLQESQDSLSSELAQL--RAQEKWHEVSFQDKEHLRLQ 694
QY 372 EAAKWEKEYGSLKDDTDDNSNPGKTDPEPKGTEAVLEAIRKNIEWLKHDKKGNKEDY 431
Db 695 DAEEMKALEQWESHREAHQKLSRLRDEIEBKQIIDEIRDLNKLQLEQEK-LSSDY 753
QY 432 DLSQMRDFINKQADAVVEKGIL--DKEE 457
Db 754 NKLKIED---QEREMKLEKLLLENDKRE 778

```

RESULT 6

```

US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IGA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Emond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-4

```

Query Match 6.9%; Score 166; DB 3; Length 1104;
Best Local Similarity 21.2%; Pred. No. 0.00057;
Matches 121; Conservative 80; Mismatches 190; Indels 181; Gaps 26;

```

QY 17 IQAFPKGQSQDKSLHNRREL-----SABRPLNEQIAEAEEDKIKTYPPENKPGQSNYSF 71
Db 139 VEKHNKP--NQOKTLKQSDTKVDLSNIDKELNHQKSQVEA-----MAEQAGITN 185
QY 72 VDNLLNLKATKEKIEKERSIRSPLDNKLNVEDVDSTKNRKLIDYDSTKSLDHF 131
Db 186 EDKDSMLKKI---EDIRKQAQ-----ADKK--EDAQVKEVREELGKLFSSTKAGLDQOI 234
QY 132 QDDPGLHQLDGTPLTAEDIVHKI-----AARIYENDRAVPDKIVSKLLNLGL 180
Db 235 QE-----HVKKET--SSEENTQKVDEHYANSLQNLAKSLEBLDKATTNEQATQVKQFL 287
QY 181 -----ITESQAHTLEDEVAEV-----LOKLISKE-----ANNYEED 211
Db 288 ENAQKLEKIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLQDLVAKSEIVREYEGK 347
QY 212 PNKPTSWTE-----NQAGKPIPEKVPMAAIQDGLAKGENDETVSNT 252
Db 348 LNQSKNLPQLBEEAHSKLVQVVEHPRKFKTSEQVTPKRVKRDLAANNQ---QK 404
QY 253 LTLTNGLERRTKYSEDNFEELQYFNFYALLKSIDSEKAKEKETLITIMTKLIDFVKM 312
Db 405 IELTVSPENITVYEGED-----VKPTVTAKS-DS-----KTTLDPSDL 441
QY 313 MVKYG-TISPEEGSYLENLD-----EMIALQTK-----NKLEKNATDN 350
Db 442 LTKYNPSVSDRISYNTKINTDNHKAETIKNKLKNSQTVTLAKADDSGNVVEKTFIT 501
QY 351 I-----SKLFPAPSEKSHETD-----STKEAAKMEKEYGSLKDS 386
Db 502 VQKKEEQVPTPEQKHSKTEQNVQBPSPKSNDRKNQQLBELIKSAQAELEKLEKAIKELMEQ 561
QY 387 TKDQNSNP--CGKTDPEKGTAEVLEAIRKNIEWLKHDKKGNKEDYDLSQMRDFINKQA 444
Db 562 -PEIPSNPEYGIQKSIWESQKEPIQEAITSFNKTIIGSSSSKYTYEHYFNKYKSHFMNYQL 620
QY 445 DAYVEKGILDKBEAE-----AIKRIYSS 467
Db 621 HAQME--ILTRKVQVMNKPDPNAEIKKIPES 650

```

RESULT 7

US-08-923-992A-6
; Sequence 6, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-6

Query Match 6.9%; Score 166; DB 3; Length 1128;
Best Local Similarity 19.3%; Pred. No. 0.00059;
Matches 119; Conservative 86; Mismatches 200; Indels 210; Gaps 23;

Qy 25 GSQKSLNRLSARPLNEQIAEEDKIKKTPPENKPGQSNVSP---VDNLLKAI 81
Db 69 GKREKQLQKQWNNLNDVDNITLSHEQKNEFKTKIDETNDSALLELENQFNETNRLHI 128
Qy 82 TEKEKIEKERSIRSPPL---DNKLNVEDVDSTKNR-----K 115
Db 129 KQHEVEKDKKAKQKQKTLKQSDTKVDLSNIDKELNHQKSPVKMAEPKGINEDKSNMLK 188
Qy 116 LIDD-----YDSTKSGLDHFKQDDPDGDLHQLDGTPLTAED 150
Db 189 KIEDIRKQAQADKEDAEVKBELGKLFSTKAGLDQE-----IHEHVKETSSEE 241
Qy 151 IVHKI-----AARIYEENDRAVDKIVSKLNLGL----- 180
Db 242 NTQKVDEHYANSLQNAQKSLLELDKATTNEQATQVKNQFLENAQKLKEMQPLIKETNVK 301
Qy 181 ITESQAHNTLEDAEV-----LQKLISKE---ANNYEEDPNKPTSWTE----- 220
Db 302 LYKAMSSELEQVEKELKHNLSANLELDLVAKSKEIVREYEGKLNQKSLPELQKLEBAHS 361
Qy 221 -----NQAGKIPKVTPEMAAIQDGLAKGENDETVSNLTITLNGLERRTKTSYDNF 271
Db 362 KKKQVVEDFRKKFTTSQVTPKVKVRDLAENNQ---QKIETVSPENITVVEGED-- 416
Qy 272 EELQYFNFYALLKSIDSEKAESKETLITIMTKTLIDFVKQMWKYG-TISPSEGVSYLEN 330

Db 417 -----VKFTVTAKS-DS-----KTTLDPSDLLTKYNPSVSDRISTNYKTN 455
Qy 331 LD-----EMIALQTK-----NKLEKNATDNISKLFPPAPSEKSHEDTST 369
Db 456 TDNHKIAEITIKNLKINESQTVTTLAKDSDGNNVVEKFTTITVQKKEKQVPTPEQKDSK 515
Qy 370 KEAA-----KWEKEYGSLKOSTKDDNPNPGKTDPEPKGT 405
Db 516 TEERKVPQEPKSNKNQQLQELIKSAQOELEKAIKELMEQ-PEIPSNP--EYGIQKSIW 572
Qy 406 EAYLEATRKNIEWLK-----HDKKNGKEDYDLSKWRDFINKQADAYVEKGLDKKEAB-- 459
Db 573 ESQKEPIQEAITSFPKLIIGDSSSKYTYTEHYFNKYKSDFMNYQLHAQME--MLTRKVVQYM 630
Qy 460 -----AIKRIYSS 467
Db 631 NKYPDNAEIKKIPES 645

RESULT 8
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cvtp036
; CURRENT APPLICATION NUMBER: US/09/595.684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 6.9%; Score 166; DB 4; Length 2662;
Best Local Similarity 21.8%; Pred. No. 0.0019;
Matches 119; Conservative 103; Mismatches 175; Indels 150; Gaps 28;

Qy 26 SQKSLNRLSARPLN---EQIAEAEEDKIKKTPPENKPGQSNYSFVDNLLKRAIT 82
Db 1101 AQEK---NHAIKKEGELSRTCRLAEVEEKLKESQQLQEQQL-----LVNQEMS 1150
Qy 83 EKKEKIEKERSIRSPDLNKLNVEDVDSTK---NRKLIDDYDSTKS-----GLDHK 130
Db 1151 EMOKKINEIENKLNKLNKELTLEHMETERLELAQKLNENVEVKSITKERKVLKELQKS 1210
Qy 131 FQDDPDGL---HQLDGTPL-TAEDI-----VH-KIARIYEENDRAVDKIVSKLNLGL 180
Db 1211 FETERDHLRGYIREIEATGLQTKKEELKIAHILKHOETIDELRRSVSEK-TAQIINTQD 1269
Qy 181 ITESQAHNTLEDAEVL---QKLIS--KEANNYEEDPNKPTSWTENQAGKIPKVTPEMAA 235
Db 1270 LEKS---HTKLOEIPVLHHEQELLPNVKYSETQETNLELLELITEQSTTK---DSTTLAR 1324
Qy 236 IQ-----DGLAK-GENDETVSNLT-----TLTNGLERRTK 264
Db 1325-1EMERLRLNEKFOESQBEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSK 1384
Qy 265 TYSEDNFEEL-----QYFPNFYALLK-----SIDSEK-AKEKE 297
Db 1385 QEOSLNNKEKDNETTKIVSEMEQPKPKDSALLRIEIMGLSKRLQSHDEMKSVAKEKD 1444

```
QY 298 TLITIMKTLIDFVKMMVKYGTISPEGVSYLENIDEMTA--LQTKNKLKLE-----KNATD 349
Db 1445 DLORLQEVL-----QSEDLQKENIKIIVAKHLETEBELKVAHCCLKEQEE 1490
QY 350 NISLFPAPSEKSHETDSTKEEAKKEKEYGSLKSTKDNPNPGKTDPEPKTE--- 406
Db 1491 TINELRVNLSEK-----STEISTIOKLEAINDKLON-----KIOEIVEKEBQLN 1535
QY 407 -AYLEAIRKNIEWLKKHDKKNKEDYDL-----SKMRDPINKQADAYVEKGILDKEEAEAI 461
Db 1536 IKQISEVQENVNELKQFKEHRKAKDSALQSIESKMLELTNRLOESQEIQIMIKEK-EEM 1594
QY 462 KRIYSSL 468
Db 1595 KRVOEAL 1601

RESULT 9
US-09-914-259-17
; Sequence 17, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-914-259-17

Query Match 6.8%; Score 163.5; DB 4; Length 956;
Best Local Similarity 20.7%; Pred. No. 0.00072;
Matches 106; Conservative 91; Mismatches 169; Indels 147; Gaps 25;

QY 25 GSQDQSLHNR-----ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLLNL- 78
Db 321 GQAKTIKTVSNVLELTAB-----EWKKYKEKEKKNKALKSVLQHLEML 367
QY 79 -----KAITEKIE-KERQSIRSPLDNKLNVED-----VDSTKNRKLIDDYDSTKSGL 127
Db 368 NNRWNGEAVPEDEQISAKDHKSL--EPCDNTPIIDNTPVWDGISAEL--EKYDEBITSL 423
QY 128 DHKFQDDPDGLHQLDGTPLTAEDIVHKAARIYEEND-----RAVFDKIVSKLLNGLIT 182
Db 424 YRLDDKDDINO-----QSQAELKIQOQMLDQDELLASTRRDYEKIQEELTRLOIEN 476
QY 183 ESQAHTLEDEVARVLOKLISKEANNYE---EDPNKPTSWTENQAGKIPKVTFMAAIDQG 239
Db 477 EA-----AKQEVKEVLQAL--BELAVNDQKSQVEDKTRANEQLTDLAQKTTTLTTTQRE 531
QY 240 LAK-----GENDETVSNTLTLNGLERR-----TKT 265
Db 532 LSQLOELSNHQKRATEILNLLKOLGEIGGIITND---VKTLDVNGVIEEFETMARL 588
QY 266 YSDNFEELQYFNFYALLKS--IDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEE 323
Db 589 YTSKMKSEKSLVNRKQLESQMDNRKNWASERELAAACOLLIQOHEAKIKSLT----- 643
QY 324 GVSYLENDEMIALQTKNLEK---NATDNISKLPAPSEKSHET--DSTKEEAAKMEK 378
Db 644 --DYMQME-----QKRRLQESQDSLSBELAKL--RAQEKHVEFSQDKEKHLTRIQ- 693
QY 379 EYGLSKDSTKDDNSNPGGKTDEPKGTEYAYLEAIRKNIEWLKKHDKKNKEDYDLSKMRD 438
Db 694 -----DAEEVYKALEQQMESHR-----BAHQKQ-----LSRLRD 722
```

```
QY 439 FINKQADAVVEKILDK-----EAEAIKRIYSSL 468
Db 723 EIEEQRIIDEIRDNLQKLEQERLSSDYNKL 755
```

RESULT 10

```
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; TITLE OF INVENTION: PORTIONS AND PRODUCTS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8991 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-714-741-32
```

```
Query Match 6.8%; Score 162; DB 4; Length 8991;
Best Local Similarity 18.2%; Pred. No. 0.02;
Matches 88; Conservative 110; Mismatches 184; Indels 102; Gaps 19;
```

```
QY 26 SODKSLHNRSLASERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLLNLKAITEKE 85
Db 2024 AQEKYADYQRIEBAKEAQESLEQQEANKDYQLKLUKYLDDGRNLNLSNVLLKMEMEEAE 2083
QY 86 KIEKERSIRSPLDNKL-----NVEDVDSTKNRKLIDDYDSTKSGLDHFKFD----- 133
Db 2084 KKDKKQEA-----EFNKIRREIVVPNPQELEMMARKSEV--VKAKESGLVKRVEAEKV 2136
QY 134 -----DPGLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVS-KLLNLGLITESQ 185
Db 2137 TEARQKLDAREAKEVLOPTRVENEVHKLXQKMNKK--KMILTSLASVAILGAGLVTSQP 2194
QY 186 AHTLEDEVAEVLQK-----LISKEANN-----YEEDPNKPTSWT 219
Db 2195 TFVRAEESPVQVEKSSLEKKYEEAKAKADTAKKDYETAKKAEDAQKYEEDQKR-----T 2250
```

QY 220 ENQAKIPE---KVTPTMA-----AIQDLAKGENDETYSNLT-LTNGLERRTK 264
 Db 2251 EERKAEASQKLDIVLVQVONAYKEYREVQNRQSKSDADYQKKLTETVDSKIEKARK 2310
 QY 265 TYS--ENFBELOVFPFYALL---KSIDSEKAEKETLITIMTKLIDFVVMVYGT 318
 Db 2311 EOODLQNNFVR-----AVVAPDPFCVGDYXRMNKKMILTSLSV-----AILGAGX 2359
 QY 319 ISPEGVSYLENDEMIALQTKNLEKATNISKLPAPSEKSHETSTKBEAAKWEK 378
 Db 2360 VTSOPTXVRAEAPQV---EKSLEKKYBEAKV-----DAAKDYDEAKKKAQAQK 2411
 QY 379 EYSLKSTKDDNSNPGKTDPEKGTKEAYLEAIRKNIEWLKHKDKGN---KEDYDLK 435
 Db 2412 KYEEDQKTEKAKAKAASEIAKATEEVQKAVLDYITAIRNHDGKTSABEAENKAK 2471
 QY 436 MRDF 439
 Db 2472 ERDY 2475

RESULT 11
 US-08-533-306A-6
 ; Sequence 6, Application US/08533306A
 ; Patent No. 5837457
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/533.306A
 ; FILING DATE: September 25, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 2115-00869COB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 641-1600
 ; TELEFAX: (810) 641-0270
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 816 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-533-306A-6

Query Match 6.6%; Score 159.5; DB 2; Length 816;
 Best Local Similarity 22.6%; Pred. No. 0.0011;
 Matches 109; Conservative 81; Mismatches 193; Indels 99; Gaps 21;
 QY 26 SQKSLHNRSLRPLNEQIARBEKIKKTPPENKPGQSNYSFV--DNMLLKAITE 83
 Db 158 SHREEMENEVSVTGMLNE--AEGKATKLAKDVASLSQLQDTQLLIQEETRQKLVSTK 215

QY 84 KEKIEKROSIR---SSPLDNKLNVEDVDSTKN-----RKLIDDYDSTKSGLDH---KF 131
 Db 216 LRQLEERNSLOQDLDEMEAKQNLERHISTLNQLSDSKKQLQDFASTVEALEEGKRF 275
 QY 132 QDDPDGLHQDGTPLTAEDIVHKIAARIYENDRAVPD-----KIVSKLNLGLITESQA 186
 Db 276 QKSIENLTQOYEKAAYDKLETKRNLQQLDLDLVVDLQNLQVSNL-----EKQ 328
 QY 187 HTLEDEVAEVLQKLSKEANNYBEDPNKPTSTWENQAGKIPKVTPTMA-AIQDLAKGEN 245
 Db 329 RKFDQLLAE--EKNIS---SKYADERDR---AEEAREKETKALSARALEEALAEKEE 379
 QY 246 DETVSNLTITNGLERRTKTYSED-----NPEELQYFPNFYALLKSIDSEKAEKETLIT 301
 Db 380 LERTNMLK---AEMEDLVSSKDDVGKNVHEL-----EKSKEALETQME 420
 QY 302 IMKTLIDFVVMVYGTISPEEGVSYLENDEMIALQTKNLEKATNISK-----L 354
 Db 421 EMKTLQEELEDELQ---ASEDAKLRLVNMQALKGQFERDQLQARDQNEEKRLQRL 476
 QY 355 FPAPSEKSHETSTKBEAAKMEKEYGSLKXSTKDDNSNPGKTDPEK--GKTEAYLEAI 412
 Db 477 HEYTELEDEDERNERALAAAKKLE-GDLKDLQADSIAKGEEAIKQLRKLQAKMDF 535
 QY 413 RKNIE-----WLKHKDKGNKEDYDLKMRDFI-----NKQADAYVEKGLDK 455
 Db 536 QRELEDARASRDEIFATAKENEKAKSLEADLMQLQEDLAAAEARAKQAD-----LEK 588
 QY 456 EE 457
 Db 589 EE 590

RESULT 12
 US-08-742-923A-6
 ; Sequence 6, Application US/08742923A
 ; Patent No. 5869611
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/742,923A
 ; FILING DATE: No. 5869611ember 1, 1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 2115-00869DVC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 641-1600
 ; TELEFAX: (810) 641-0270
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 816 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; US-08-742-923A-6
;
; Query Match
; Best Local Similarity 6.6%; Score 159.5; DB 2; Length 816;
; Matches 109; Conservative 81; Mismatches 193; Indels 99; Gaps 21;
;
; 26 SQDKSLHRELNSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITKE 83
; 158 SHREEMENEVESVTGMLNE--AEGKAIKLAKDVASLSQQLDTQELLQEEETROKLNVS TK 215
;
; 84 KEKTEKERSQSTR---SSPLDNKLNVEDVDSTKN-----RKLIDDYDSTKSGLDH---KF 131
; 216 LRQLEERENSLQDQDEEMAEKQNLERHISTLNQLSDSKKKLQDFASTVEALEEGKKRF 275
;
; 132 QDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQA 186
; 276 QKEIENLTQQYEKAAAYDKLETKNRLOQELDDLVLDNQRLVSNL-----EKKQ 328
;
; 187 HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPTMA-AIQDGLAKGEN 245
; 329 RKFPQLLAE--EKNIS---SKYADERDR---AEAAREKETKALSARALEEALKEE 379
;
; 246 DETVSNLTNLNGLRRTKTYSED-----NFEELOYFPNFYALLKSIDSEKAKEKETLIT 301
; 380 LERTNKM LK---AEMEDLVSSKDDVGKQNVHEL-----EKSRALETQME 420
;
; 302 IMKTLIDPVMVMVYGTISPEEGVSYLENDEMIALQTKNLEKNATDNISK-----L 354
; 421 EMKTLQELSEDELQ-----ASEDAKLRLVNMQALKGQFERDLQARDEQNEKRRQLQRL 476
;
; 355 FPAFSEKSHETDSTKEBAAKMEKEYGSLKSDTKDDNSNPGGKTDEPK--GKTEAYLEAI 412
; 477 HEYTELEDERNERALAAAKKLE--GDLKDLQELQADSIAIKGREBAIKQLRKLQAKMDF 535
;
; 413 RKNIE-----WLKXHDKGNKEDYDLSKMRDFI-----NKQADAYVEKGILDK 455
; 536 QRELEDAARSDEIFATAKENEKKAKEADLMQLQEDLAAAEARAKQAD-----LEK 588
;
; 456 EE 457
; 589 EE 590
;
; RESULT 13
; US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533.306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-533-306A-4
;
; Query Match
; Best Local Similarity 6.6%; Score 158.5; DB 2; Length 885;
; Matches 109; Conservative 79; Mismatches 190; Indels 99; Gaps 21;
;
; 31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITKEKIE 88
; 232 LQNEVESVTGMLNE--AEGKAIKLAKDVASLSQQLDTQELLQEEETROKLNVS TKLQ 289
;
; 89 KERQIR---SSPLDNKLNVEDVDSTKN-----RKLIDDYDSTKSGLDH---KFQDD 136
; 290 BERNLQDQDEEMAEKQNLERHISTLNQLSDSKKKLQDFASTVEALEEGKKRFQKEIE 349
;
; 137 GLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQAHTLED 191
; 350 NLTYQYEKAAAYDKLETKNRLOQELDDLVLDNQRLVSNL-----EKKQKRFQ 402
;
; 192 EVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPTMA-AIQDGLAKGENDET 250
; 403 LLAE--EKNIS---SKYADERDR---AEAAREKETKALSARALEEALKEELETRN 453
;
; 251 NTLTITNLNGLRRTKTYSED-----NFEELOYFPNFYALLKSIDSEKAKEKETLITIM 306
; 454 KMLK---AEMEDLVSSKDDVGKQNVHEL-----EKSRALETQMEEMKTQ 494
;
; 307 IDFKVMVMVYGTISPEEGVSYLENDEMIALQTKNLEKNATDNISK-----LFPAPS 359
; 495 LEELEDELQ-----ASEDAKLRLVNMQALKGQFERDLQARDEQNEKRRQLQRLHEYET 550
;
; 360 EKSHEETDSTKEBAAKMEKEYGSLKSDTKDDNSNPGGKTDEPK--GKTEAYLEAIRKNIE 417
; 551 ELEDERNERALAAAKKLE--GDLKDLQELQADSIAIKGREBAIKQLRKLQAKMDFQRELE 609
;
; 418 -----WLKXHDKGNKEDYDLSKMRDFI-----NKQADAYVEKGILDKEE 457
; 610 DARASRDEIFATAKENEKKAKEADLMQLQEDLAAAEARAKQAD-----LEKEE 659
;
; RESULT 14
; US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/742,923A
 FILING DATE: No. 5869611 member 1, 1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Smith, DeAnn F.
 REGISTRATION NUMBER: 36683
 REFERENCE/DOCKET NUMBER: 2115-00869DVC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (810) 641-1600
 TELEFAX: (810) 641-0270
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 885 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-742-923A-4

Query Match 6.6%; Score 158.5; DB 2; Length 885;
 Best Local Similarity 22.9%; Pred. No. 0.0015;
 Matches 109; Conservative 79; Mismatches 190; Indels 99; Gaps 21;

Qy 31 LHNRELSAERPLNQGIAAEEDKIKKTPPENKPGQSNYSFV--DNLLKKAITEKEKIE 88
 Db 232 LQNEVESVTGLNE--AEGKAIKLAKDVASLSQLODTQELLQETROKLNVTSLRQLE 289

Qy 89 KEROSIR--SSPLDNKLVNVEDVSTKN-----RKLIDDYDSTKGLDH--KFODDPD 136
 Db 290 EERNLQDQDEEMAKONLHRHSTLNLIQSDSKKLQDPFVTEALBEGKGRFQKEIE 349

Qy 137 GLHOLDGTPLTAEDIVHKIAAIIYEENDRAVFD-----KIVSKLLNLGLITESQAHTLED 191
 Db 350 NLTQYEEKAAAYDKLETKNRLQELDDLVDLDNQQLVSNL-----EKKQKFDQ 402

Qy 192 EVAEVLQKLISKEANNVEEDPNKPTSWTENOAGKIPEKVTMA--AIQDGLAKGENDETVS 250
 Db 403 LLAE--EKVNI--SKYADERDR---AAEAAREKETKALSARALBEALEAKEELERTN 453

Qy 251 NTLTLTNGLERRTKYSIED---NFEELQYFPNFYALLKSIDSEKAKEKTLITIMKTL 306
 Db 454 KVLK-----AEMEDLVSSKDDVGKVNHEL-----EKSRALETQMEEMKTQ 494

Qy 307 IDFKVMVKYGTISPGEVSYLENLDEMIALQTKNLEKNATDNISK-----LFPAPS 359
 Db 495 LEELEDELQ---ASEDAKLELVNMQALKQGFEDLQARDEQNEEKRRQLOQLHHEYET 550

Qy 360 EKSHEETDSTKEAAKMEKEYGSLKSDTKDSDNSPGKTDDEPK--GKTEAYLEAIRKNIE 417
 Db 551 ELEDERNERALAAAANKLE--GDLKDLQELQADSIAIKGREBAIKQLRLQAKMDQFORELE 609

Qy 418 -----WLKHKDKGNKEDYDLSKMRDFI-----NKQADAVVEKGILDKKE 457
 Db 610 DARASRDEIFATAKENEKKALEADLMQLEDLAAAARARKQAD-----LEKKE 659

RESULT 15
 US-09-914-259-12
 ; Sequence 12, Application US/09914259
 ; Patent No. 6495336
 ; GENERAL INFORMATION:
 ; APPLICANT: Makowski, Lee
 ; APPLICANT: Hyman, Paul
 ; APPLICANT: Williams, Mark
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
 ; FILE REFERENCE: 8471-010-999
 ; CURRENT APPLICATION NUMBER: US/09/914,259
 ; CURRENT FILING DATE: 2000-11-21
 ; NUMBER OF SEQ ID NOS: 180
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12

LENGTH: 1087
 TYPE: PRT
 ORGANISM: Oryctolagus cuniculus
 US-09-914-259-12

Query Match 6.5%; Score 156.5; DB 4; Length 1087;
 Best Local Similarity 19.3%; Pred. No. 0.0028;
 Matches 103; Conservative 83; Mismatches 180; Indels 167; Gaps 22;

Qy 31 LHNRELSAERPLNQEIAAEEDKIKKTPPENKPGQSNYSFVNDLNLKKAITEKEKIEKE 90
 Db 304 IHEKEVEIDR-LNEQIIKLOQ-----LKITDNKVIEEK 337

Qy 91 QSISSPLDNKLVNVEDVST-KNR-----KLIDDYDSTKGLDHKQDDPDGLHQL 141
 Db 338 NELIRDLEAQIECLMSQOERVNRNREBEIEQLNEVIEKQLQELANIQKTSVDPSSLSL- 396

Qy 142 DGTPLTAEDIVHKIAAIIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVA----- 194
 Db 397 -----EADSLKHQL-----DKVIAEKLAL-----EHQVETTNEMAVTKNVLK 434

Qy 195 -----EVLQKLIS--KEANNVEEDPNKPTSWTENOAGKIPEKVTMAAI--ODGLAKG 243
 Db 435 ETNFKMQLTQELCSLKREREKMERIQSVPEKSVNMSVGLSKDKKPEMDLIPTEDALAQL 494

Qy 244 E-----NDETVSNTL-TLTNGLERTKTYSDNFEELQYFPNFYALLKSIDSEKAKE 295
 Db 495 ETQTQLRSSESSKVSLSLETLLQLESTVSTKDLTLTQY-----KQIQDMRQGR 547

Qy 296 KET-----LITIMKTL-----IDFKVMVKYGTISPGE----- 324
 Db 548 SETEMLQTKIVSLQKVLEEKVAAALVSQVLEAVQEVYKLCADKPAVSSDPARTEVPCLS 607

Qy 325 -----VSYLENL-----DEMIALQTKNLEK-NATDNISKLF 355
 Db 608 QLAGNTMESDVSAITWRISELSQVEMHSSLISEKEQVIAEKNALEKKEKLOELQKLV 667

Qy 356 PAPSEKSHETDSTKEAAKMEKEYGSLKSDTKDSDNSNPG-----GKTDEPKGKTBAYL 409
 Db 668 QDSETKQRE-----RERQSRHLGDLGVLESTTSEESGVFGELEALRAESAAPKGE LANYK 722

Qy 410 EATRKNIEWLKKHDKGNKEDYDLSKMRDFIKQADA---YVEKGILDKKEAE 459
 Db 723 ELAEKLEELLVKETNMASLPKLSHVDRDQLEAEDKLSHFSEKE--DKTEVQ 773

Search completed: January 2, 2004, 12:36:20
 Job time : 22 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	2399	100.0	468	11	US-09-931-836-67		Sequence 67, Appl
2	2399	100.0	468	12	US-10-035-977-67		Sequence 67, Appl
3	2399	100.0	468	12	US-10-063-735-150		Sequence 150, Appl
4	2399	100.0	468	12	US-10-137-870-90		Sequence 90, Appl
5	2399	100.0	468	12	US-10-140-018-90		Sequence 90, Appl
6	2399	100.0	468	12	US-10-140-021-90		Sequence 90, Appl
7	2399	100.0	468	12	US-10-140-274-90		Sequence 90, Appl
8	2399	100.0	468	12	US-10-140-471-90		Sequence 90, Appl
9	2399	100.0	468	12	US-10-140-807-90		Sequence 90, Appl
10	2399	100.0	468	12	US-10-140-922-90		Sequence 90, Appl
11	2399	100.0	468	12	US-10-140-924-90		Sequence 90, Appl
12	2399	100.0	468	12	US-10-140-926-90		Sequence 90, Appl
13	2399	100.0	468	12	US-10-141-698-90		Sequence 90, Appl
14	2399	100.0	468	12	US-10-141-703-90		Sequence 90, Appl
15	2399	100.0	468	12	US-10-141-704-90		Sequence 90, Appl

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: P3030R1C10
CURRENT APPLICATION NUMBER: US/10/035,977
CURRENT FILING DATE: 2001-12-26
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/112514
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113300
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113430
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114140
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115552
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/125774
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125826
PRIOR FILING DATE: 1999-03-24
PRIOR APPLICATION NUMBER: 60/127035
PRIOR FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/127706
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/129122
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 60/130359
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131272
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/132371
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132379
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132383
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138166
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/144791
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/146970
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380142
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/644848
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 09/747259
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/816744
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/854208
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/854280

```

; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 67
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-035-977-67

Query Match      100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8,7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy   1  MGFLGTGTVILVLVLP IQAPKPGSQD KSLHNRELSAERPLNQIAEAEEDKKIKTYPP 60
Db   1  MGFLGTGTVILVLVLP IQAPKPGSQD KSLHNRELSAERPLNQIAEAEEDKKIKTYPP 60

Qy   61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKRQSIRSPDLNKLNVEDVDSTKNRKLIDDY 120
Db   61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKRQSIRSPDLNKLNVEDVDSTKNRKLIDDY 120

Qy   121 DSTKSGLDHFKPD DPGHLQJLGDPTLTAEADIVHKIIARIYEENDRAVPDKIVSKLLNLGL 180
Db   121 DSTKSGLDHFKPD DPGHLQJLGDPTLTAEADIVHKIARIYEENDRAVPDKIVSKLLNLGL 180

Qy   181 ITESQAHTLEDEVAEVLOKLI SKENNYEEDPNKPTSWTENQAGKIPEKVTPMAAIODGL 240
Db   181 ITESQAHTLEDEVAEVLOKLI SKENNYEEDPNKPTSWTENQAGKIPEKVTPMAAIODGL 240

Qy   241 AKGENDETVSNTLTTLTNGLERRTKYTSGEDNFEELOYFPNFWALLKSIDSEKEAKEKETLI 300
Db   241 AKGENDETVSNTLTTLTNGLERRTKYTSGEDNFEELOYFPNFWALLKSIDSEKEAKEKETLI 300

Qy   301 TIMKTLLDFVKMMVKYGTISPEGVSYLENLDEMIALQTNRKNEKNATDNISKLFPAESE 360
Db   301 TIMKTLLDFVKMMVKYGTISPEGVSYLENLDEMIALQTNRKNEKNATDNISKLFPAESE 360

```

QY 361 KSHEETDSTKEAAKMEKEYGSLKSTKDDNSNPGKTDPEKGTAEYLAIRKNIEWLK 420
DB 361 KSHEETDSTKEAAKMEKEYGSLKSTKDDNSNPGKTDPEKGTAEYLAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
DB 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 3

US-10-063-735-150
; Sequence 150, Application US/10063735
; Publication No. US2003013882A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,735
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 150
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-735-150

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGTWILVLPIQAFPKPGSQSKLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
DB 1 MGFLGTGTWILVLVLPIQAFPKPGSQSKLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
QY 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKQIRSPDLNKNVEDVDSTKRLIDDDY 120
DB 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKQIRSPDLNKNVEDVDSTKRLIDDDY 120
QY 121 DSTKSGLDHFKFQDDPDGLHOLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
DB 121 DSTKSGLDHFKFQDDPDGLHOLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKIPKVTMAAIQDGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKIPKVTMAAIQDGL 240
QY 241 AKGENDETNSNTLTNLGLRRTKTYSEDNFELOYPNFPYALLKSIDSEKAEKETLI 300
DB 241 AKGENDETNSNTLTNLGLRRTKTYSEDNFELOYPNFPYALLKSIDSEKAEKETLI 300
QY 301 TIMKTLIDFVKMVKYGTISPEGVSYLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
DB 301 TIMKTLIDFVKMVKYGTISPEGVSYLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
QY 361 KSHEETDSTKEAAKMEKEYGSLKSTKDDNSNPGKTDPEKGTAEYLAIRKNIEWLK 420
DB 361 KSHEETDSTKEAAKMEKEYGSLKSTKDDNSNPGKTDPEKGTAEYLAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
DB 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 4

US-10-137-870-90
; Sequence 90, Application US/10137870
; Publication No. US2003013883A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tuma, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C155
; CURRENT APPLICATION NUMBER: US/10/137,870
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-870-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGTWILVLPIQAFPKPGSQSKLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
DB 1 MGFLGTGTWILVLVLPIQAFPKPGSQSKLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
QY 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKQIRSPDLNKNVEDVDSTKRLIDDDY 120
DB 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKQIRSPDLNKNVEDVDSTKRLIDDDY 120
QY 121 DSTKSGLDHFKFQDDPDGLHOLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
DB 121 DSTKSGLDHFKFQDDPDGLHOLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKIPKVTMAAIQDGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKIPKVTMAAIQDGL 240
QY 241 AKGENDETNSNTLTNLGLRRTKTYSEDNFELOYPNFPYALLKSIDSEKAEKETLI 300
DB 241 AKGENDETNSNTLTNLGLRRTKTYSEDNFELOYPNFPYALLKSIDSEKAEKETLI 300
QY 301 TIMKTLIDFVKMVKYGTISPEGVSYLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
DB 301 TIMKTLIDFVKMVKYGTISPEGVSYLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
QY 361 KSHEETDSTKEAAKMEKEYGSLKSTKDDNSNPGKTDPEKGTAEYLAIRKNIEWLK 420
DB 361 KSHEETDSTKEAAKMEKEYGSLKSTKDDNSNPGKTDPEKGTAEYLAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
DB 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESIST 5

```

US-10-140-018-90
; Sequence 90, Application US/10140018
; Publication No. US2003013885A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRADE
; TITLE OF INVENTION: ACIDS ENCODING
; FILE REFERENCE: P330R1C158
; CURRENT APPLICATION NUMBER: US/10/140-018-90
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Patent
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-018-90

```

Query Match	100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity	100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	MGFLGTGWIILVLPQTQAPPKPGSGQKSLHNRELSAERPLNEQITAEAEEDKI KKTYP 60
DB	
1	MGFLGTGWIILVLPQTQAPPKPGSGQKSLHNRELSAERPLNEQITAEAEEDKI KKTYP 60
QY 61	ENKPGQSNYSFVDNINLLKATTEKEKIEKERQSRSSPLDNKUNVEDVSTKRNKLIDDY 120
DB	
61	ENKPGQSNYSFVDNINLLKATTEKEKIEKERQSRSSPLDNKUNVEDVSTKRNKLIDDY 120
QY 121	DSTKSLGDHKKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
DB	
121	DSTKSLGDHKKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
QY 181	ITPSQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTPMMAIQDGL 240
DB	
181	ITPSQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVTPMMAIQDGL 240
QY 241	AKGENDETVSNTLTLTNGLERRTKTYSEDNFELOYFPNFIYALLKSIDSEKAKEKETLI 300
DB	
241	AKGENDETVSNTLTLTNGLERRTKTYSEDNFELOYFPNFIYALLKSIDSEKAKEKETLI 300
QY 301	TIIMKTLIDFVKMWVKYGTISPGEGVSVLENLDEMIALQTKNLEKNATDNISKLFPAPSE 360
DB	
301	TIIMKTLIDFVKMWVKYGTISPGEGVSVLENLDEMIALQTKNLEKNATDNISKLFPAPSE 360
QY 361	KSHEETDSTKEEAAKMEKEVGS�KDSPTKDNNSPGGKTDEPKGTEAYLEAIRKNIEWLK 420
DB	
361	KSHEETDSTKEEAAKMEKEVGS�KDSPTKDNNSPGGKTDEPKGTEAYLEAIRKNIEWLK 420
QY 421	KHDKKGNKEDYDLISKMRDFINKQADAYVEKGILDKSEABAIKRIYSSL 468
DB	
421	KHDKKGNKEDYDLISKMRDFINKQADAYVEKGILDKSEABAIKRIYSSL 468

RESULT 6

U.S.-10-140-021-90

```

US-10-104-021-90
/ Sequence 90, Application US/10140021
/ Publication No. US20030138886A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Ellen
/ APPLICANT: Filvaroff, Luc
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerstein, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin J.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECURED AND TR
/ TITLE OF INVENTION: ACIDS ENCODING
/ FILE REFERENCE: P330R1C167
/ CURRENT APPLICATION NUMBER: US/10/1
/ CURRENT FILING DATE: 2002-05-06
/ Prior Application removed - See Pat
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 90
/ LENGTH: 468
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-140-021-90

```

[illegible]

RESULT 7

RESULT /
US-10-140-274-90

; Sequence 90, Application US/10140274
; Publication No. US20030143674A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C161

; CURRENT APPLICATION NUMBER: US/10/140,274

; CURRENT FILING DATE: 2002-05-06

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 90

; LENGTH: 468

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-140-274-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8,7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGFLGTGWTWLVLP	1QAPFKPGSQD	KS	LHN	RELS	SAER	PLNE	QIAE	AEED	KIK	TYPP	60
Db	1	MGFLGTGWTWLVLP	1QAPFKPGSQD	KS	LHN	RELS	SAER	PLNE	QIAE	AEED	KIK	TYPP	60
QY	61	ENKPGQSNYSFV	DNLNLLK	KAITE	KEK	ERQ	SRSS	PLDN	KNLN	VED	VDST	KNR	KLI
Db	61	ENKPGQSNYSFV	DNLNLLK	KAITE	KEK	ERQ	SRSS	PLDN	KNLN	VED	VDST	KNR	KLI
QY	121	DSTKSGLDH	KFDQDDP	GLH	QD	GTPL	TAED	I	VHK	IAA	RIE	END	RAV
Db	121	DSTKSGLDH	KFDQDDP	GLH	QD	GTPL	TAED	I	VHK	IAA	RIE	END	RAV
QY	181	ITESQAHTLE	DEVAEVL	QKLI	SK	EANN	YED	PNK	PTS	WTEN	OAG	KI	PEK
Db	181	ITESQAHTLE	DEVAEVL	QKLI	SK	EANN	YED	PNK	PTS	WTEN	OAG	KI	PEK
QY	241	AKGENDET	VNTLT	LT	NG	LERR	TK	TY	SE	DN	FE	ELQ	YFP
Db	241	AKGENDET	VNTLT	LT	NG	LERR	TK	TY	SE	DN	FE	ELQ	YFP
QY	301	TIMKTLIDF	VVMV	KYGT	IS	PE	GV	SY	LEN	DEM	IAL	Q	T
Db	301	TIMKTLIDF	VVMV	KYGT	IS	PE	GV	SY	LEN	DEM	IAL	Q	T
QY	361	KSHEETD	STKE	AAK	ME	KEY	GS	L	K	D	ST	K	D
Db	361	KSHEETD	STKE	AAK	ME	KEY	GS	L	K	D	ST	K	D
QY	421	KHDKGNK	EDY	D	L	S	K	M	R	D	F	I	N
Db	421	KHDKGNK	EDY	D	L	S	K	M	R	D	F	I	N

RESULT 8
US-10-140-471-90
; Sequence 90, Application US/10140471

; Publication No. US20030138887A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C163

; CURRENT APPLICATION NUMBER: US/10/140,471

; CURRENT FILING DATE: 2002-05-06

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 90

; LENGTH: 468

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-140-471-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8,7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGFLGTGWTWLVLP	1QAPFKPGSQD	KS	LHN	RELS	SAER	PLNE	QIAE	AEED	KIK	TYPP	60
Db	1	MGFLGTGWTWLVLP	1QAPFKPGSQD	KS	LHN	RELS	SAER	PLNE	QIAE	AEED	KIK	TYPP	60
QY	61	ENKPGQSNYSFV	DNLNLLK	KAITE	KEK	ERQ	SRSS	PLDN	KNLN	VED	VDST	KNR	KLI
Db	61	ENKPGQSNYSFV	DNLNLLK	KAITE	KEK	ERQ	SRSS	PLDN	KNLN	VED	VDST	KNR	KLI
QY	121	DSTKSGLDH	KFDQDDP	GLH	QD	GTPL	TAED	I	VHK	IAA	RIE	END	RAV
Db	121	DSTKSGLDH	KFDQDDP	GLH	QD	GTPL	TAED	I	VHK	IAA	RIE	END	RAV
QY	181	ITESQAHTLE	DEVAEVL	QKLI	SK	EANN	YED	PNK	PTS	WTEN	OAG	KI	PEK
Db	181	ITESQAHTLE	DEVAEVL	QKLI	SK	EANN	YED	PNK	PTS	WTEN	OAG	KI	PEK
QY	241	AKGENDET	VNTLT	LT	NG	LERR	TK	TY	SE	DN	FE	ELQ	YFP
Db	241	AKGENDET	VNTLT	LT	NG	LERR	TK	TY	SE	DN	FE	ELQ	YFP
QY	301	TIMKTLIDF	VVMV	KYGT	IS	PE	GV	SY	LEN	DEM	IAL	Q	T
Db	301	TIMKTLIDF	VVMV	KYGT	IS	PE	GV	SY	LEN	DEM	IAL	Q	T
QY	361	KSHEETD	STKE	AAK	ME	KEY	GS	L	K	D	ST	K	D
Db	361	KSHEETD	STKE	AAK	ME	KEY	GS	L	K	D	ST	K	D
QY	421	KHDKGNK	EDY	D	L	S	K	M	R	D	F	I	N
Db	421	KHDKGNK	EDY	D	L	S	K	M	R	D	F	I	N

RESULT 9
US-10-140-807-90
; Sequence 90, Application US/10140807
; Publication No. US20030134354A1

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C174
; CURRENT APPLICATION NUMBER: US/10/140,807
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-807-90

Query Match      100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 MGFLGTGTLVLVLPIQAFPKPGSQDKSLHNRLSAAERPLNEQIAAEEDKIKKTYPP 60
Db 1 MGFLGTGTLVLVLPIQAFPKPGSQDKSLHNRLSAAERPLNEQIAAEEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLLKALTEKEKIEKERQSISSPLDNKLNVEDVDSTKNRKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKALTEKEKIEKERQSISSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIARIIYENDRAVFDKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIARIIYENDRAVFDKIVSKLLNLGL 180

Qy 181 ITESQAHTLEDEVAEVLQKLSKANNYEEDPNKPTSWTENOAGKIPKVTPTMAIQDGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKANNYEEDPNKPTSWTENOAGKIPKVTPTMAIQDGL 240

Qy 241 AKGNDETVSNTLTLTNGLERRTKITYSDNFEEQLQYFNFYALLKSIDSEKEAKEKETLI 300
Db 241 AKGNDETVSNTLTLTNGLERRTKITYSDNFEEQLQYFNFYALLKSIDSEKEAKEKETLI 300

Qy 301 TIMKTLIDFVKQMWKYGTISPEEGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKQMWKYGTISPEEGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360

Qy 361 KSHETSTKEEAAKMEKEYGSLKDSKTDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
Db 361 KSHETSTKEEAAKMEKEYGSLKDSKTDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy 421 KHKDKGNKEDYDLGSKMRDFPNKQADAYVEKGIIDKEEAETKRIYSSL 468
Db 421 KHKDKGNKEDYDLGSKMRDFPNKQADAYVEKGIIDKEEAETKRIYSSL 468

RESULT 10
US-10-140-922-90
; Sequence 90, Application US/10140922
; Publication No. US2003013889A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary B.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: F3330R1C179
; FILE REFERENCE: F3330R1C179
; CURRENT APPLICATION NUMBER: US/10/140,922
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-922-90

Query Match      100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 MGFGTGTWLVLVPIQAPKPGSQDSKSLHNRELSAERPLNEQIAEABEDKIKKTYPP 60
Db 1 MGFGTGTWLVLVPIQAPKPGSQDSKSLHNRELSAERPLNEQIAEABEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLKKAITEKEKIEKRSQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLKKAITEKEKIEKRSQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy 121 DSTKSGLDHFKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy 181 ITTSSQHTLDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAQKIPKVTTPMAAIQDGL 240
Db 181 ITTSSQHTLDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAQKIPKVTTPMAAIQDGL 240

Qy 241 AKGENDSTVNTLTLTNGLRRTTKTYSSEDNPEELQYFPNFVALLKSIDSEKAEKETLI 300
Db 241 AKGENDSTVNTLTLTNGLRRTTKTYSSEDNPEELQYFPNFVALLKSIDSEKAEKETLI 300

Qy 301 TIMKTLIDFVKVMVKYGTISPEEGVSYLENDLDEMIALQTKNLEKNATDNISKLPFPAPSE 360
Db 301 TIMKTLIDFVKVMVKYGTISPEEGVSYLENDLDEMIALQTKNLEKNATDNISKLPFPAPSE 360

Qy 361 KSHETDSTKEEAAKMEKYGSLKSDTKDONSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
Db 361 KSHETDSTKEEAAKMEKYGSLKSDTKDONSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy 421 KHKDKGNKEDYDLSKMRDFTNKQADAVVEKGLDKEEAETKRIYSSL 468
Db 421 KHKDKGNKEDYDLSKMRDFTNKQADAVVEKGLDKEEAETKRIYSSL 468

RESULT 11
US-10-140-924-90
; Sequence 90, Application US/10140924
; Publication No. US20030134355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```

RESINT 17

RESULT 11
US-10-140-924-90

US-10-140-924-90
; Sequence 90, Application US/10140924

; sequence 3V, Application US/101
; Publication No. US20030134355A1; FUDICATION NO. US20
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.


```

; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C206
; CURRENT APPLICATION NUMBER: US/10/141,698
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-141-698-90

```

Query Match	100.0%	Score	2399	DB 12	Length	468				
Best Local Similarity	100.0%	Pred. No.	8.7e-150							
Matches	468	Conservative	0	Mismatches	0	Indels	0	Gaps	0	
Qy	1	MGFLGTGTWLVLP	IQAPPKPGSQD	KSLHNREL	SAERPLN	EQIAEABED	KIKKTYP	60		
Db	1	MGFLGTGTWLVLP	IQAPPKPGSQD	KSLHNREL	SAERPLN	EQIAEABED	KIKKTYP	60		
Qy	61	ENKPGQSNYSFV	DNILNLLKA	ITEKEKIE	KERQSI	RRSPLDN	KLNVEDV	DSSTNKRKL	IDDY 120	
Db	61	ENKPGQSNYSFV	DNILNLLKA	ITEKEKIE	KERQSI	RRSPLDN	KLNVEDV	DSSTNKRKL	IDDY 120	
Qy	121	DSYKSGLDHKE	FDDPDG	LHQDGTPL	TAEDIV	HKIAARI	YEENDRA	VPDKIVSK	LLNGL 180	
Db	121	DSYKSGLDHKE	FDDPDG	LHQDGTPL	TAEDIV	HKIAARI	YEENDRA	VPDKIVSK	LLNGL 180	
Qy	181	ITBSQAHTLDE	VAEVLQK	LISKANN	YEEDPN	KPTSWT	ENQAQGI	PEKVTPMAA	IODGL 240	
Db	181	ITBSQAHTLDE	VAEVLQK	LISKANN	YEEDPN	KPTSWT	ENQAQGI	PEKVTPMAA	IODGL 240	
Qy	241	AKGENDTVSNT	LTLTNG	LERRTK	YSEDN	PEELQY	PNFYALL	KSIDSEKAE	KEKTLI 300	
Db	241	AKGENDTVSNT	LTLTNG	LERRTK	YSEDN	PEELQY	PNFYALL	KSIDSEKAE	KEKTLI 300	
Qy	301	TIWKTLIDFVK	MMVKYGT	I	SPEEGV	SYLEN	DEMIALQ	TNNKLEKN	ATDNI	SKLFPAPSE 360
Db	301	TIWKTLIDFVK	MMVKYGT	I	SPEEGV	SYLEN	DEMIALQ	TNNKLEKN	ATDNI	SKLFPAPSE 360
Qy	361	KSHEETDST	KEEAAKME	KEYGSL	KDSTK	DNSNPG	KTDDEPK	GKTEAY	LEAIRN	NIWLK 420
Db	361	KSHEETDST	KEEAAKME	KEYGSL	KDSTK	DNSNPG	KTDDEPK	GKTEAY	LEAIRN	NIWLK 420
Qy	421	KHKDKGNK	EYD	YLSKMR	DFIN	KQADAY	VEKGIL	DKBAAE	AIKRI	YSSSL 468
Db	421	KHKDKGNK	EYD	YLSKMR	DFIN	KQADAY	VEKGIL	DKBAAE	AIKRI	YSSSL 468

RESULT 14
US-10-141-702-90
; Sequence 90, Application US/10141702
; Publication No. US20030134358A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc

```

?  APPLICANT:  Filvaroff, Ellen
?  APPLICANT:  Gao, Wei-Qiang
?  APPLICANT:  Gerritsen, Mary E.
?  APPLICANT:  Goddard, Audrey
?  APPLICANT:  Godowski, Paul J.
?  APPLICANT:  Gurney, Austin L.
?  APPLICANT:  Sherwood, Steven
?  APPLICANT:  Smith, Victoria
?  APPLICANT:  Stewart, Timothy A.
?  APPLICANT:  Tumas, Daniel
?  APPLICANT:  Watanabe, Colin K
?  APPLICANT:  Wood, William
?  APPLICANT:  Zhang, Zemin
?  TITLE OF INVENTION:  SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
?  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
?  FILE REFERENCE:  P3330R1C208
?  CURRENT APPLICATION NUMBER:  US/10/141.702
?  CURRENT FILING DATE:  2002-05-08
?  Prior Application removed - See Palm or File wrapper
?  NUMBER OF SEQ ID NOS: 550
?  SEQ ID NO 90
?  LENGTH: 468
?  TYPE: PRT
?  ORGANISM: Homo Sapien
?  US-10-141-702-90

```

Query Match	100.0%;	Score 2399;	DB 12;	Length 468;
Best Local Similarity	100.0%;	Pred. No. 8,7e-150;		
Matches 468;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGFLGTGTTILVLVLPLOAPKPGCGSQDKSLHNHRELSAERPLNBOIAFAEEDKIKKTYPP	60	
Db	1	MGFLGTGTTILVLVLPLOAPKPGCGSQDKSLHNHRELSAERPLNBOIAFAEEDKIKKTYPP	60	
Qy	61	ENKPGQSNYSFVDNLNLKAI TEKEKTEKEROSSRSPLDNKNLVNEDVDS TKNRKLIIDDY	120	
Db	61	ENKPGQSNYSFVDNLNLKAI TEKEKTEKEROSSRSPLDNKNLVNEDVDS TKNRKLIIDDY	120	
Qy	121	DSTKSGLDHDKFQDDPDLGHLQDGTPLTAEDIVHKHIAARIYEENDRAVDFKIVSKLLNLGL	180	
Db	121	DSTKSGLDHDKFQDDPDLGHLQDGTPLTAEDIVHKHIAARIYEENDRAVDFKIVSKLLNLGL	180	
Qy	181	ITTSQAHTLDEVAEVLQKLI SKEANNYEEDPNKPTSWTENQAQKIPKVTTPMAAIQDGL	240	
Db	181	ITTSQAHTLDEVAEVLQKLI SKEANNYEEDPNKPTSWTENQAQKIPKVTTPMAAIQDGL	240	
Qy	241	AKGENDETVSNTLTLTNLGLERRTKTYSSEDNFEELQYFPNFPYALLKSIDSEKEAKEKETLI	300	
Db	241	AKGENDETVSNTLTLTNLGLERRTKTYSSEDNFEELQYFPNFPYALLKSIDSEKEAKEKETLI	300	
Qy	301	TI MKTLIDFVKMMVKYGTI SP EEGVSYLENIDEMALQTKNLEKATDNISKLFPAPSE	360	
Db	301	TI MKTLIDFVKMMVKYGTI SP EEGVSYLENIDEMALQTKNLEKATDNISKLFPAPSE	360	
Qy	361	KSHEETDSTKEEAAKMEKYGSLKSDTKDQNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK	420	
Db	361	KSHEETDSTKEEAAKMEKYGSLKSDTKDQNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK	420	
Qy	421	KHPDKGNKEDYDLSKMRDFINQOADA VVEKGILDKBEAAEAIKRIYSSL	468	
Db	421	KHPDKGNKEDYDLSKMRDFINQOADA VVEKGILDKBEAAEAIKRIYSSL	468	

RESULT 15
US-10-141-704-90
; Sequence 90, Application US/10141704
; Publication No. US20030134359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen

APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330RIC209
CURRENT APPLICATION NUMBER: US/10/141,704
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 90
LENGTH: 468
TYPE: PRT
ORGANISM: Homo Sapien
US-10-141-704-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGWILVLPVLPQAPKPGSQDSKSHNELSAERPLNEQIAEAEDKIKKTYPP 60
DB 1 MGFLGTGWILVLPVLPQAPKPGSQDSKSHNELSAERPLNEQIAEAEDKIKKTYPP 60
QY 61 ENKPGQSNYSFVDNLNLLKATEKEKIERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
DB 61 ENKPGQSNYSFVDNLNLLKATEKEKIERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
QY 121 DSTKSGLDHKKFQDDPGDGLHQLDGTPLTAEDIVHKIARIYEENDRAVFDKIVSKLNLGL 180
DB 121 DSTKSGLDHKKFQDDPGDGLHQLDGTPLTAEDIVHKIARIYEENDRAVFDKIVSKLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKPEKVTNMAAIODGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKPEKVTNMAAIODGL 240
QY 241 AKGENDETVENTLTNGLERRTKTYSEDNFEELQYFPNPFYALLKSIDSEKAKEKETLI 300
DB 241 AKGENDETVENTLTNGLERRTKTYSEDNFEELQYFPNPFYALLKSIDSEKAKEKETLI 300
QY 301 TIMKTLIDFVKMMVKYGTISPESGVSVLENLDEMIALQTKNKNATDNISKLPAPSE 360
DB 301 TIMKTLIDFVKMMVKYGTISPESGVSVLENLDEMIALQTKNKNATDNISKLPAPSE 360
QY 361 KSHHEETDSTKEAAKMEYKSLKSDKDNNSPGGKTDEPKGTEAYLEAIRKNIEWLK 420
DB 361 KSHHEETDSTKEAAKMEYKSLKSDKDNNSPGGKTDEPKGTEAYLEAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
DB 421 KHDKGNKEDYDLKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

Search completed: January 2, 2004, 12:37:45
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 20 Seconds
(without alignments)
2250.347 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTLVLVLPIQAP.....EKGILDKKEBAIKRIYSSL 468
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1957	81.6	533	2 A37180	chromogranin/secre
2	196	8.2	2139	2 T18296	myosin heavy chain
3	191	8.0	2464	1 Q8MSPL	microtubule-associ
4	186.5	7.8	2116	2 A26655	myosin heavy chain
5	184	7.7	665	2 B71609	hypothetical prote
6	183.5	7.6	1790	2 S67593	transport protein
7	180.5	7.5	2364	2 A56577	microtubule-associ
8	179.5	7.5	1526	2 A45605	mature-parasite-in
9	179	7.5	853	2 T51505	hypothetical prote
10	179	7.5	1005	2 A64485	hypothetical prote
11	179	7.5	1558	2 B71603	RESA-H3 antigen PF
12	179	7.5	2269	2 T28677	rhoxy protein -
13	178.5	7.4	1392	2 A43336	microtubule-vesicl
14	178.5	7.4	1427	2 S22695	restin - human
15	177	7.4	1871	2 D36796	probable heat shoc
16	172.5	7.2	1134	2 A60234	IGA Fc receptor pr
17	172.5	7.2	1164	1 FCSOAG	IGA Fc receptor pr
18	172.5	7.2	1979	2 C71622	hypothetical prote
19	171.5	7.1	3488	2 T34178	hypothetical prote
20	170	7.1	1804	2 T34518	neatin - golden ha
21	169	7.0	1272	2 C30593	hypothetical prote
22	168.5	7.0	746	2 T47237	myosin II heavy ch
23	168.5	7.0	926	2 T24923	hypothetical prote
24	168	7.0	2401	2 T28676	rhoxy protein -
25	167	7.0	976	2 T01553	hypothetical prote
26	166.5	6.9	1875	2 S38173	myosin-like protei
27	166.5	6.9	5327	2 T13564	microtubule-associ
28	166	6.9	991	2 H86168	hypothetical prote
29	166	6.9	1192	2 A71623	probable secreted

30 166 6.9 2663 1 S28261 centromere protein
31 164.5 6.9 719 2 A81358 hypothetical prote
32 164.5 6.9 911 2 S51441 hypothetical prote
33 164.5 6.9 1127 2 T28317 ORF MSV156 hypothe
34 164.5 6.9 1269 2 F84730 probable myosin he
35 164 6.8 1909 2 A45592 liver stage antige
36 163 6.8 1252 2 B42771 reticulocyte-bindin
37 163 6.8 2510 2 T28160 hypothetical prote
38 162.5 6.8 1738 2 T14867 interaptin - slime
39 162.5 6.8 2385 2 A32491 myosin heavy chain
40 162.5 6.8 2411 2 B32491 myosin heavy chain
41 161.5 6.7 1302 1 JC6009 surface-located me
42 161.5 6.7 1992 1 S02771 myosin heavy chain
43 161 6.7 852 2 D72230 conserved hypothet
44 160 6.7 1354 2 S74244 serine/threonine-s
45 159.5 6.6 821 2 S67087 hypothetical prote

ALIGNMENTS

RESULT 1

A37180
Chromogranin/secretogranin-like vesicle protein precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 30-Sep-1993
C;Accession: A37180
R;Otter, H.P.; Battenberg, E.F.; Tsou, A.P.; Bloom, F.E.; Sutcliffe, J.G.
J. Neurosci. 10, 3135-3147, 1990
A;Title: IB1075: a brain- and pituitary-specific mRNA that encodes a novel chromogranin/
A;Reference number: A37180; MUID:90376160; PMID:2204688
A;Accession: A37180
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-533 <OTT>

Query Match 81.6%; Score 1957; DB 2; Length 533;
Best Local Similarity 86.7%; Pred. No. 3.5e-87;
Matches 383; Conservative 21; Mismatches 38; Indels 0; Gaps 0;

Qy 16 PIQAFKPGGSDKSLHNRELSAERPLNEQIAEAEEDKIKKTPPENKPGQSNYSFVDNL 75
Db 63 PMSFPKPGSDKSLHNRELSAERPLNEQIAEAEADKIKKTPSEKSPERNFSSVDNL 122
Qy 76 NLLKAITEKEKIEKQSISSPLDNKLVNEDVDSTKNRKLIDDYDSTKGLDHFQDDP 135
Db 123 NLLKAITEKETVEKAKQSISSPFDNVLNVDADSTKNRKLTDYDSTKGLDKVQDDP 182
Qy 136 DGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLNLGLITESQAHTLEDEVAE 195
Db 183 DGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLNLGLITESQAHTLEDEVAE 242
Qy 196 VLOKLISKEANNYEEDPNKPTSTENQAGKIPKVTFPMAAIODGLAKGENDETSTNLTLL 255
Db 243 ALQKLISKEANNYEAEPEKPTSTENQDGKIPKVTFVAATQDFTNRENDTSTNLTLL 302
Qy 256 TNGLERTKYSDNFEELQYFNFVALLKSIDSEKAEKETLITIMKTILDFVKVMVK 315
Db 303 SNGLERTNPHRDDFEELQYFNFVALLTSIDSEKAEKETLITIMKTILDFVKVMVK 362
Qy 316 YGTISPEEGVSYLENLDEMIALQTKNLEKNADNTNISKLPAPSEKSHETDSTKEBAK 375
Db 363 YGTISPEEGVSYLENLDETALQTKNLEKNTTDSKSLFPAPPEKSHETDSTKEBAK 422
Qy 376 MEKEYSLKSDTDDNSNPGKTDPEKGTAEVLEAIRKNIEWLKKDKGNKEDYDLK 435
Db 423 MEKEYSLKSDTDDNSNPGKTDPEKGTAEVLEAIRKNIEWLKKDKGNKEDYDLK 482
Qy 436 MRDFINKQADAYVEKGILDKKE 457
Db 483 MRDFINQADAYVEKGILLIRKK 504

T18296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T18296
R:Guillen, N
submitted to the EMBL Data Library, February 1997
A:Reference number: Z18865
A:Accession: T18296
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-references: EMBL:L03534; NID:gi850912; PID:gi850913; FIDN:AAB48065.1
C:Genetics:
C:Gene: mhca
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 8.2%; Score 196; DB 2; Length 2139;
Best Local Similarity 21.0%; Pred. No. 0.1;
Matches 107; Conservative 108; Mismatches 191; Indels 104; Gaps 23;

QY 20 FPKPGSGSDSKLHNRELSAERPLNQLQIAEAEDKIKTKYPPEN-----KPGQSNYSFV 72
Db 859 FEEEGKKDKIED-----LKKQLAE---EIKGREAENALASATATGLEAKIQ 906
QY 73 D-----NLNLLKAITEKIEKIERQSISSPLDNKLNVEDVD- 109
Db 907 DLEDKISELSKLSAAELDKQELNKLKIENLEEDKEELKETIDNLKGDLDKSKLKGEDLEV 966
QY 110 -----STKQKLIIDYDSTKSGLDHKFQDDPDGHLQDGTPLTAEDIVHKIAARIY 160
Db 967 EITELNSQINTLN-ATVDKDKTTAEMOESIDEKEDEITKLKGDIKLLEEKODL----- 1020
QY 161 EENDR---AVFDKIVSKLLNGLITESAQHTLEDEVAEVLQKLISKEANNYEDPNKPT 216
Db 1021 -EQDRAVYSATKDDIAKLNKITI-----ECDAKDEIAKLQEQLEDEE--NKNKDLTNEL 1073
QY 217 SWTNQAGKIPKVPMPAAI-----ODGLAKG-ENDETVSNITLTNG-LERRTKTYS 267
Db 1074 QQTQLKLGTEKSLAAQAAVATKASDERDTLSQNLNEKLTNNLTKTADLEKKISGLK 1133
QY 268 EDNFEELQYFNFVALLKSID-----SEKAKEKETITIMKTLIDFV-KMMVKYGTISP 321
Db 1134 QD-YEDLEDDQN-----KIEGDLRNAQRKIKELDDEITKGADVSVQLQKQKEY----- 1181
QY 322 EBGVSYLENDEMIALQTKNKLKNAATDNISKLPAPSEKSHETDTKKEAAK----ME 377
Db 1182 ESQIAKQWEEKEALGNDVKNK-EKTIKE--KELEIQSLQEKLDTEVEKEDEAKKKKEIE 1238
QY 378 KEYSGLKDSKTDKDDNSNGGKTDEPKGKTEAYLEAIRKNIEWLKHKDKGNKEDYDL-SKM 436
Db 1239 KEMKALQEE-KENVESKNSSTKDKKLEDLNLDKDTQKLLDDMTADNEKLAKAKDLQEAQL 1297
QY 437 RDTFNKQADAVVEKGIIDKEEAERAIKIYS 466
Db 1298 NEVDQNHKAVADAELLNKKKAQSDKELNS 1327

RESULT 3
QRMSP1
N;Alternate names: microtubule-associated protein MAP1B - mouse
C;Species: Mus musculus (house mouse)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
C:Accession: S07549; S43387; A33645
R;Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A;Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A;Reference number: A33645; MUID:90094539; PMID:2480963
A;Accession: S07549
A:Molecule type: mRNA

Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
 A>Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum
 A;Reference number: A26655; MUID:87092266; PMID:3540939
 A;Accession: A26655
 A;Molecule type: DNA
 A;Residues: 1-2116 <WAR>
 A;Cross-references: GB:M11938; NID:g167834; PIDN:AAA33227.1; PID:g167835
 R;DeLozanne, A.; Lewis, M.; Spudis, J.A.; Levinwand, L.A.
 Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
 A;Reference number: A24728; MUID:86016788; PMID:3901008
 A;Accession: A24728
 A;Molecule type: mRNA
 A;Residues: 2035-2116
 R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
 FEBS Lett. 227, 71-75, 1988
 A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain
 A;Reference number: S00250; MUID:88112226; PMID:2828113
 A;Accession: S00250
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1734-1893 <WAG>
 C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
 C;Superfamily: myosin heavy chain; myosin motor domain homology
 C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phosphate binding site
 F:1-818/Domain: Globular head <HED>
 F:819-747/Domain: myosin motor domain homology <MMOT>
 F:179-186/Region: nucleotide-binding motif A (P-loop)
 F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 7.8%; Score 186.5; DB 2; Length 2116;
 Best Local Similarity 21.6%; Pred. No. 0.29;
 Matches 108; Conservative 86; Mismatches 205; Indels 101; Gaps 17;

Qy 26 SQDKS--LHNRELAEPLNQIAEAEDKIKTYPENKPGQSNYSFVDNLNLKAIITE 83
 Db 1023 TKDKSELLRQKKLEELKQVQALAEATAKLAQEAANKLQGEYTEL-NEKFNSEVTA 1081

Qy 84 KEIEKERQSTRSS--PLDNKLVNEDVDSTNKRKLIDYDSTKSGLDHKKFQDDPGLHLQ 141
 Db 1082 RSNVEKSKTLESQVAVNNELDEE---KQNR---DALEKKKALDAMLEMKDQLEST 1134

Qy 142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLNLGLITESQAHTLDEVAEVLKLI 201
 Db 1135 GGEKKSLYDL-----VKQSDMEALRNQISELQSTIAKLEKTLSTLGEVAVRLQGE 1188

Qy 202 SKS--ANNVEDPNKPTSWTNOAGKIPKVT-----PMAAQDGLAKEN 245
 Db 1189 AEQLAKSNVEKQKKVVELDLQKSLAEATAAKQALDKLKKLEQELSEVQTQLSEANN 1248

Qy 246 D-----ETVSNLTITLNGLERTKTYSEDN----FEELYFPNFYALLKSIDSE 290
 Db 1249 KVNNSDSTNKHLETFSNNKLELEAEQAKQALEKRLGLESELKH-----VNEQLEEE 1302

Qy 291 KEAKEK-----ETLITIMKTLDIFKVMVKVGTITSPSEGVSYLE----- 329
 Db 1303 KQKQESNEKRKVDLEKEVSELSKQDIEEVASKAVTEAKNKKSELSDELIRKQYADVSSR 1362

Qy 330 --NLDEMLATQKN-KLEKNATDNISKLFPAPSEKSHETD-----STKEAAKMEKEY 380
 Db 1363 DKSVEQLTKLAQKEELNTEAEAGQLDRAERKKAFLDEAVKNLEETAKKVAE 1422

Qy 381 GSKLDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNIEWLKHKDKGNKEDYDLSKMRDFI 440
 Db 1423 KAMKKAETDYRSTRKSELDDAKNVSSEQVQIKRLNEE-----LSLSRLV 1467

Qy 441 NKQADAYVEKGILOKEEABA 460
 Db 1468 -EADERCNSAIKAKTKAES 1486

RESULT 5
 B71609
 hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum
 C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C;Accession: B71609
 R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ;; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
 Science 282, 1126-1132, 1998
 A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A;Reference number: A71600; MUID:99021743; PMID:9804551
 A;Accession: B71609
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-665 <GAR>
 A;Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g3845245
 A;Experimental source: clone 3D7
 C;Genetics:
 A;Gene: PFB0680w

Query Match 7.7%; Score 184; DB 2; Length 665;
 Best Local Similarity 22.0%; Pred. No. 0.089;
 Matches 104; Conservative 76; Mismatches 182; Indels 110; Gaps 17;

Qy 25 GSODKSLHNRELAEPLNQIAEAEDKIKTYPENKPGQSNYSFVDNLNLKAIITEK 84
 Db 160 GKQDISNNAE--NKQDVKGVELEKKEEKISDDHKVEENKKS-----DD 205

Qy 85 EKIEKERQSTRSSPLDNKLN---VEDVDSTNKRKLIDYDSTKSGLDHKKFQDDPGLHLQ 140
 Db 206 HKVEENKKSDDHKVEENKKSDDHKIEEVKVEEHEEDEDKKEKSENKKNKDNKDEND 265

Qy 141 LDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLNLGLITESQAHTLDEVAEVLKLI 200
 Db 266 EDNDEISDEDDVDVEEDKNENDIDDDK-----KETDKTHLEEEENIEEKE 314

Qy 201 IS-KEANNYEDPNKPTSW-TENQAGKIPKVTPEMAAIQDGLAKGENDETNTLTITNG 258
 Db 315 FSDKQKNGKNDYKESKQTEKESKDIEK-----EKSKQKEKESK----- 356

Qy 259 LERRTKTYSEDNFEELYFPNFYALLKSIDSEKAEKETLITIMKTLDIFVGMVKYGT 318
 Db 357 -KQKEKEKQDKKE-----KSKDIEKE-KEKDKDIEKESK-DTAKEKEKDKD 402

Qy 319 ISPEGVSYLENLDEMLATQKNKLEKATNTNISKLPAPSEKSH-----E 364
 Db 403 IEKESK-----KQMEKLNKQNDKDKDDNEKK-----KNDQDIDHDDNDDEMBEIE 451

Qy 365 ETSTKEAAKME-----KEYGLKDSYK-DDNSNPGGKTDEPKGTEAYLEAI 412
 Db 452 ENDEDEDEDMENKKKKKGKNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 511

Qy 413 RKNIEWLKHKDKGNKEDYDLSKMRDF-----INKQADAYVEKGILO 454
 Db 512 NENNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 563

RESULT 6
 S67593
 transport protein US01 - yeast (Saccharomyces cerevisiae)
 N;Alternate names: protein D2552; protein YDL058w
 C;Species: Saccharomyces cerevisiae
 C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
 A;Accession: S67593; A38455; S30782
 R;Bloeker, H.; Brandt, P.
 submitted to the Protein Sequence Database, July 1996
 A;Reference number: S67587
 A;Accession: S67593
 A;Molecule type: DNA
 A;Residues: 1-1790 <BLO>
 A;Cross-references: EMBL:274106; NID:g1431058; PID:e253003; PID:g1431059; MIPS:YDL058w
 A;Experimental source: strain S288C
 R;Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamaeaki, M.
 J. Cell Biol. 113, 245-260, 1991
 A;Title: A cytoskeleton-related gene, US01, is required for intracellular protein transp
 A;Reference number: A38455; MUID:91185402; PMID:2010462

A:Accession: A38455
A:Molecule type: DNA
A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NA>
A:Cross-references: GB:X54378, NID:64777; PIDN:CAA3253.1; PID:94778
A:Note: the authors translated the codon ACT for residue 768 as Ile
R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A:Description: An integrin analogue in *Saccharomyces cerevisiae*.
A:Reference number: S30782
A:Accession: S30782
A:Molecule type: DNA
A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580, 'S'
A:Cross-references: EMBL:L03188
C:Genetics:
A:Gene: SGD:USO1; INT1
A:Cross-references: SGD:S0002216; MIPS:YDL058W
A:Map position: 4L
C:Keywords: coiled coil; transmembrane protein
F:326-342/Domain: transmembrane #status predicted <TM1>
F:394-410/Domain: transmembrane #status predicted <TM2>
F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 7.6%; Score 183.5; DB 2; Length 1790;
Best Local Similarity 20.3%; Pred. No. 0.33;
Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;

QY 29 KSLHN--RELAEPLNQLAEAEEDKIKYPPENKPGQSNYSFVNDNLKAI---T 82
Db 965 KSLANNYKDMQAE--NESLIKAVEE-----SKVESIQSLNQLKIDMSQ 1008

QY 83 EKEKIEKERSQIRSPDNKUNVEDVSTQVKLI-----DYDSTKSLDHFQD--- 133
Db 1009 EKENFQIERGSIENIEQLKTKTIDLEQTKBEIISKSDSSKDEYESQISLLKEKLEATT 1068

QY 134 -DPDLGDLGDTPLTAEDIVHKIARIYEENDRAVDFKIVSKLNLGLITESQAH----- 187
Db 1069 ANDENVNKISLTKTRELEAEALAA--YKNLKNELETJLSTSEKALKEVENEHEHLKEEK 1126

QY 188 -TLDEVAEVLQKLIKEAN-----NYEED-PNKPTSWTENQAGKIPEK 229
Db 1127 IQLEKEATETKQLNSLRANLESLEKEHDLAQLKKYEQIANKGRQYNE-EISQINDE 1185

QY 230 VTPMAAIDGLAKGENDETVNTLTNLGERRT-----KTYSDN-- 270
Db 1186 ITSTQOENESIKK-KNDELGEVAMKSTSEBOSNLKKSFDALNQLIKELKKKNETNEA 1244

QY 271 -----FEELOVFPNF-----YALLKSTDSSEKAK 294
Db 1245 SLLESIKSVESETVKIKELQDECNFKEVSELEDKIKASEDKNKLQK-ESEKIKE 1303

QY 295 EKETLIITMKTLDIFVKQMWK----- 315
Db 1304 ELDAKTTTELKIQLEKITNLKAKEKSESELSRLKKTSSERKNAERQLEKLNKNEIQIKNQ 1363

QY 316 -----YGTISPE--EGSVYLENLDLMIALQTKNKLKNATDNISKLPAPSEK 361
Db 1364 AFEKERKLLNEGSTTTOEYSEKINILE--DEIRLQNEENELKAKBIDNTRSELEKVSLS 1421

QY 362 SHEETOSTKEAAAMEKYEYGLSKDS-TKDDNSNPGKTDPE-----KGTAYLEAIRK 414
Db 1422 NDELLBEKQNTIKSLQDEILSYKDKITRNDKLLSTERNKRDLESLEKQLRAAQESKAK 1481

QY 415 NIEWLAKHDKGKNKEDYDLSKMRDFINKQADAYVEKGILD-KEEFAIRK 463
Db 1482 VEELGLKLEESSEKAELEKSEKEM-KKLESTIESNETELKSMETIRK 1530

RESULT 7
A56577
microtubule-associated protein MAP 1B - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C:Accession: A56577

R:Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
Eur. J. Cell Biol. 57, 66-74, 1992

A:Title: Identification of two distinct microtubule binding domains on recombinant rat M
A:Reference number: A56577; MUID:92347374; PMID:1639092

A:Accession: A56577

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2364 <ZAU>

A:Cross-references: GB:X60550

A:Experimental source: brain

A:Note: nucleotide sequence not given; conceptual translation not complete

C:Superfamily: microtubule-associated protein MAP1B

Query Match 7.5%; Score 180.5; DB 2; Length 2364;
Best Local Similarity 22.7%; Pred. No. 0.65;
Matches 120; Conservative 77; Mismatches 200; Indels 131; Gaps 22;

QY 22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIK-KYPPENKPGQSNYSFVNDNLKLA 80
Db 454 KPLSSKSVKESKEAPEATKASQVEKTPKVESKEKIVKKDKPGK-----VESKPS 505

QY 81 ITEKIEKERSQIRSPDNKUNVEDVSTQVKLIIDYDSTKSLDHFQDDP--DGL 138
Db 506 VTEKEVPSKEEQSPVKAEEVAEKAATESPKVTQKDKVVKKEIKTKP--EKKKEKPKKEVA 563

QY 139 HOLDGDTPLTAEDIVHKIARIYEENDRAVDFKIVSKLNLGLITESQAH--LEDEVAEVL 197
Db 564 KKEDKTPPLKDKDEKPKK-----EBAKKEIKKEIKKEKELKKEVKKETPLKDAKKEV- 615

QY 198 QXLIKEKANNVEEDPNK-----PTSWTENQAG-----KIKEKVTPTMAA- 235
Db 616 KKDEKKEVKEKEPKKEIKKIDKIKSTPLSDTKPKPALPKPVAKKEETKEPIAAG 675

QY 236 -IQD-GLAK-----GENDETVSNTL-----TLTNGLERRTK 264
Db 676 KLQDKGKVKVKKEGKTTTAAATAVGTAAVAAAAGVAASGPAKELAEBSLSWSPEDLTK 735

QY 265 TVSDNFEELQVFPNFYALLKSIDSEKAE-----KETLIITMKTLDIFVKQMW 314
Db 736 DFEELEKAEIDVADIKPQLSLIDEKELKETEPEGEAYVIQKETEVS----- 782

QY 315 KYGTISPEGVSVYLENLDLMIALQTKNKL---EKNATDNISKL---FPAPSEKSHSE 365
Db 783 KGSAESPDGEGITTEGECE--CEQTPPELEPVEKQGVDDIEKFEDEGAGFESEAGDYE 840

QY 366 TDSITKEEAAKMEKYEYGLSKDSTKDDNSNPGKTDPEKQTEA--YLBAIRNI----- 416
Db 841 EKAETEAEEPEEDG---EDNVSGSASKHSPTDEEETAKAEADVHIKEKRESVASGDRA 897

QY 417 -----EWLKKHDKGKNKEDYDLSKMRDFINKQADAYVEKGILDKKEAE 459
Db 898 EEDWDEALEKGEAQSEEGEEEEE-----DKAEDAREEDHEPDKTEAE 940

RESULT 8

A56505

mature-parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium

C:Species: Plasmodium falciparum

C>Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: A45605; A54517

R:Coppel, R.L.

Mol. Biochem. Parasitol. 50, 335-347, 1992

A:Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human ery

A:Reference number: A45605; MUID:92158014; PMID:1741020

A:Accession: A45605

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1526 <COP>

A:Experimental source: Papua New Guinean isolate FC27

A:Note: sequence extracted from NCBI backbone (NCBI:83648, NCBI:83656)

R:Coppel, R.L.; Culvenor, J.G.; Bianco, A.E.; Crewe, P.E.; Stahl, H.D.; Brown, G.V.; J

Mol. Biochem. Parasitol. 20, 265-277, 1986

A:Title: Variable antigen associated with the surface of erythrocytes infected with matu

A:Reference number: A54517; MUID:87014571; PMID:3531849
A:Accession: A54517
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 222-443, 'K', 445 <C02>
A:Cross-references: GB:M15319; NID:g160060; PID:g552170
C:Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology
C:Keywords: surface antigen; tandem repeat

Query Match 7.5%; Score 179.5; DB 2; Length 1526;
Best Local Similarity 18.7%; Pred. No. 0.42;
Matches 107; Conservative 101; Mismatches 182; Indels 181; Gaps 24;

```
QY 25 GSQKSLHNRSL--AERPLNEQTAABEDKIKTYPPENKPGQSN-----68
DB 400 GASSENSEDPKLTQEENGTGKESSEETKDDK-----PEENEKADNKKSKKKKSPFQM 454
QY 69 --YSPVDNLN-----LKAITEKIEKERSIRSSPLDNKLVNEDVDSTNKRKLIDDYD 121
DB 455 LGCNPLCNKNIETDDEETLVVKODAKKKHFLREA-----NTEKNDNEKKKLLGEDG 508
QY 122 -----STKSGLDHKFQDDPDGL-----HOLDGTPL 146
DB 509 KEDVKKNDEQKVLGEGDKEDVKNDEQKVLGEGDKEDVKNDEGKKKGVISEK 568
QY 147 TAEDIVHKIAARI-----YEENDRAVFDKIVSKLLNLGLITESQAHTLEDEV-- 193
DB 569 TQKEIKERKVRVKKCKKKVKKGKENDTEGNDKVGPEIIEBVKEEIKKQVEDGIKE 628
QY 194 -----AEVLQKLSKEANNYEDPNKPTSWTENQAGKIPKVT-----MAAI 236
DB 629 NDTGNDKVGPEIIEBVKEEIKKQVEDGKENDTEGNDKVGPEIIEBVKEEIKKQV 688
QY 237 QDGLAG--ENDETNSNTLTITNGLERTK-----TYSEDNP-----EELQY 276
DB 689 EEGIKENDTEGNDKVGPEIIEBVKEEIKKQVEGVKENDTESKDKVIGOEIITEVK- 747
QY 277 PNFYALKSIDSEKAKEKTLTI-----MTLIDFVQMVKYGTISPGEV--SYL 328
DB 748 -----KETQBEKNGKENILEIKDIVIGQEVIIIEVVKVIK--KVEKGIKENHT 796
QY 329 ENLD-----EMALQTKNKLKNATDNISKLPAPSEK-----SHEETDSTKEA 373
DB 797 ESKDKVIGOEIIEBVKEEIEKQVEGVKENDTESKDKVIGVNEGVNBPENKDKV 856
QY 374 AKMEKEYGSLKDDKDDNSNPGKTDPEKGTAEYLEAIRKNIEWLKHKDKKNGKEDYDL 433
DB 857 TKQEK--VKEVK-----EVKKK-----VKRV--KRNKNKRNKONVI 891
QY 434 SK--MRDPINKQADAYVEKGILDKKEEAPAK 462
DB 892 GKEIMKEDVNEKTDANKDKTEIQEKEKEVK 922
```

RESULT 9
T51505
hypothetical protein F5E19_70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51505
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A:Reference number: 225394
A:Accession: T51505
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <SAT>
A:Cross-references: EMBL:AL391147
A:Experimental source: cultivar Columbia; BAC clone F5E19
C:Genetics:
A:Map position: 5
A:Introns: 6/2; 79/3
A:Note: F5E19_70

Query Match 7.5%; Score 179; DB 2; Length 853;
Best Local Similarity 20.9%; Pred. No. 0.21;
Matches 102; Conservative 81; Mismatches 206; Indels 100; Gaps 19;

```
QY 16 PIQAFPKPGSQKSLHNRSLAER-----PLNEQIAEAEDKIKTYPPEN-----62
DB 45 PSTTTPHSRSLDRSSPNSKSSVERRSPKLTTPPEKSOARVAAVKGTESPTTTTLRSLQIK 104
QY 63 ---KPGQSNYSFVDNLNLLKAIATEKIEKERSIRSSPLDNKLVNEDVDSTNKRKLIDD 119
DB 105 EDLKKANERISSLEK-DKAKALDELKQAKAEQV-TLKLDDALKAQ--KHVEENSIEK 160
QY 120 YDSTKSGLDHKFQDDPDGL-----HOLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
DB 161 FOAVEAGIE-AVQNNBEELKKELETVKVQHASDAALVA---VRQELEKINEEL-AAAFD 215
QY 170 KIVSKLLNLGLITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKIPK 229
DB 216 -AKSKAL-----SQA-----EDASKTAIEIHAERKVDILSE 244
QY 230 VTPMAATQDGLAKGENDETVSNLTITNGLERRTKYSEDNFEELQYFPNFYALKKIDS 289
DB 245 LTRLKALLD-----STREKTAISDNEMVAKLEDEIVVLKRDES 283
QY 290 -----EKEAKEKETLITIMKTLIDFVQMVKYGTISPGEVSYLENDEMIALQTKNLEK 345
DB 284 ARGFEAEVKEKEMIVKLVNDELAAKMAESNAHSLSNMQSKAKELSEQ--LEEANKLER 341
QY 346 NAT---DNISKLFPAPSEKSHETDST--KEAAAKMEKEYGSLKDDKDDNSNPGKTD 399
DB 342 SASVLSVNMVKLEGNDKLHDTETITDLKERIVTLTETTVAKQKEDLEVSEQR--GSVE 400
QY 400 EPKGKTEAYLEAIRKNIEWLKHKDKKNGKEDYDLSKMRDFINKQADAYVEKGILDKKEAE 459
DB 401 EEVSKNEKEVEKLSLELTVKEKNRALKKEQATSRVQLSEBKSLSDLESSKEEE 460
QY 460 AIKRIYSSL 468
DB 461 KSKKAMESL 469
```

RESULT 10
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: A64465
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, C.;
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
A:Reference number: A64300; MUID:96337999; PMID:868087
A:Accession: A64465
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1005 <BL>
A:Cross-references: GB:U67572; GB:L77111; NID:g1591958; PIDN:AAB99331.1; PID:g1591962; TJ
C:Genetics:
A:Map position: REV1273394-1270377
C:Superfamily: hypothetical protein MJ1322

Query Match 7.5%; Score 179; DB 2; Length 1005;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 116; Conservative 80; Mismatches 164; Indels 160; Gaps 27;

```
QY 52 DKIKTKYPPENKPGQSNYSFVDNLNLLKA-ITEKEKEKERQSRSSPLDNKLVNEDVD 109
DB 159 DEFEKCY--QKGEIVKEVEKLERIEGELNKENYEK-----LQNKMSQLESE-- 205
QY 110 STNKRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
```

```

Db 206 --KMKLMEINDKLNK- IKKSEFIEKLPFNENKLLYKFKLNK-----EERKALEL 257
QY 170 KIVS-KLNLGLITSSQAH-----TLEDEVAEVLQKLSKEANNVEEDPNK 214
Db 258 KNQELKILEYDLNTVVEARETNRRHDEYKYSVLVDEIRKIESRL--RELKSHYEDYLK 315
QY 215 PISWTENQAGKIPEKVTWMAAIQDGLAGKENDETVSTNTLTNGLERRTKITYSDNPEEL 274
Db 316 LTKQLEIIKGDIEK-----LKEFINKSKYRDDIDNLDTLNKKI-----KDEIERV 360
QY 275 QYFPNFPYALLKSIDSEKAEK-----ETLITIMKTLDFVMMVKYGTISPE 322
Db 361 ETIKOLLEBELKNLEIEIKYKRICBECKEYVEKYLEBEKAVEYNKLTLEVITL-- 417
QY 323 EGVSVLENDEMIALQTKNLEKNATD--NISKLPAPS-----EKSHEEDTSTK-- 370
Db 418 -----LQEKSKTEKNINDLETRINKLEKNIDIESIENSUKIEEEKKV 463
QY 371 -----EEAKMEKEVGSLSKSTK-----DINSNPGGK-----TDEPK 405
Db 464 LENQEKIELNKLGEINSEIKLKLDELKEVEGKPLCKTPIDENKQWELINQHT 523
QY 406 E--AYLE-----AIRKNIEWLKKH-----DKGNKEDVDL----- 433
Db 524 QLNKKYTELEINKKIRIEKIDIEKKEIDKEENLKLTKLYLEKQSQIEELELKNY 583
QY 434 SKMRDFINQADAYVEK-----ILD--KEBAEAIKRIYS 466
Db 584 KEQLDEINKKISNVYNGKPVDEILEDIKSOLNKFKNFYN 623

RESULT 11
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71603
R:Gardner, M.J.; Fettel, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Pierce, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: B71603
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1558 <GAR>
A:Cross-references: GB:AE001424; GB:AE001362; NID:93845307; PIDN:AACT71972.1; PID:9384530
A:Experimental source: clone 307
C:Genetics:
A:Gene: PFB0915w

```

```

Query Match 7.5%; Score 179; DB 2; Length 1558;
Best Local Similarity 21.5%; Pred. No. 0.45;
Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;

QY 27 QDKSLHRELAE--RPLAQIAEAEEDKIKKYPENKPGQSN-----YSFVDNL-- 75
Db 725 EENAVSNENVAENLEKLNTEVNTVLDKVEETVEISGESLENEMDMKAFSEIFDNVKG 784
QY 76 ---NLK-----AITEKEKIEKERSIRSSPLDNKLN-----VEDVDSTK-- 112
Db 785 IQENLLTGMFRSITSIVIOSEKVDL--NENVVSSILDNIEKNKGLNKLNISSSTEGV 843
QY 113 -----NRKLDDVD-----STKSGLDHKQDDPDGLHQDGTPLTAEDI- 151
Db 844 QETVTEHVQNVYVDVDPAMKQDQFGLINEAGGLKEMFNLED--VFKSESVDITVEBK 902
QY 152 -----VHKIARIYEENDRAVEDKIVSKLNLGLITQSQAHTLEDEVAEVLQKLSKE 204
Db 903 DEPQVEKEKETSIIIEENNIIVD-----VLEEKEKEDTKMDADAEIESIS 951
QY 205 ANNYEEDPNKPTSWTENQAGKIPEKVTWMAAIQDGLAGKENDETVSTNTLTNGLERRTK 264

```

```

Db 952 SDSKEE-----TES-----IKQEKQVSLVVEEQDNDMDSEVKVLELKNMBEELMK 999
QY 265 TYSEDN-----PEELQYFPNFYA--LLKSIDSEKAEK-----EKETLITIMK 304
Db 1000 DAVEINDITSKLIEETQELNEVEADLIKMEKLEKALSSEDSKEIIOAKODTLEKVIIE 1059
QY 305 -----TLIDFVMMVKYGTISPE--EGVSVLENDEMIALQTK--NKLEKNATDNISK 354
Db 1060 EEHDITTTTLDV---VELKDVVEDKIEKVSIDLKQLEEDILKEVKEIKELSEILEDYKEL 1116
QY 355 FPAPSEKSEHEEDTSTKEBAKMEKEVGSLSKSTKDNNSNPGKTDPEKGTAYLEAIRK 414
Db 1117 KITIEDILKEKEIEKDHFEKFEBAEIKOLEADILKEVSVLESEVEBEKKLEVEVHE--LKE 1175
QY 415 NIEWLKHKDK--KGNKEDYDLSKMRDFINQADAY---VEKGILDK-----EBA 458
Db 1176 EVEHITSGDAHIKGLEED--DLEEVDDLKGSILDKGDMELGDMDKESLEDTVAKLGERV 1234
QY 459 EAIKRIYSS 467
Db 1235 ESLKQVLISS 1243

RESULT 12
T28677
rhostry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhostry protein of Plasmodium yoelii.
A:Reference number: Z20508; MUID:95021522; PMID:7935623
A:Accession: T28677
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEE>
A:Cross-references: EMBL:L27838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

```

```

Query Match 7.5%; Score 179; DB 2; Length 2269;
Best Local Similarity 21.1%; Pred. No. 0.73;
Matches 113; Conservative 85; Mismatches 193; Indels 144; Gaps 23;

QY 13 LVLIQIAPPKGGQDQKSLHRELAEAEEDKIKKYPENKPGQSNYSFV 72
Db 7 LILPLMNQFKGLNESMIKLNKSGILRKYTISNQI---KNKLVNSTYPEGREG-----FT 57
QY 73 DNMLN-----LKAITEKEK-----TEKERSIRSSPLDN-----KLN- 104
Db 58 SSLELAKSWEKTKLETITELTKSNEETVRLKEIREULFKYLDDEAEERKYLEGLKLELNK 117
QY 105 -----VEDVDST-----KNRKLIDD-----YDSTKSGLDH 129
Db 118 KIKDIIAKIEYVKNVTELKKEIKKNAYIDELANQSPYKVTGYIENKNTIYNTIKSYFDQ 177
QY 130 KFQDDPGLHQDGTPLTAEDIHVHKAARIYEENDRAVDFKIVSKLNLGLIT-ESQAH 188
Db 178 IYEGDIDTFYN--ELSSIVKEDPIDDDIEDTKLENLSKIDNVYDKIQKMBIETVKSHLNN 236
QY 189 LE-----DEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIPEKVT 231
Db 237 IETNNKLPNTILEKKIYVIDEISKELNKMLEDPFNKEKELSNKISDY--DKRRELSEYKS 295

```



```

QY 232 PMAAIQDGLAKGENDET-VSNLTILTNGLERRTKTYSEDNFEELQYFPNFVALLKSIDSE 290
Db 296 KMLEIRNH-----NSQTNVDNTE-----EAKQNYDKSN-EHMTTIPITNEDEISKIISE 345
QY 291 KEAKEKETLITIMKTLPIDFVMMVKYGTISPPEGVSYLENDEMIALOTKMKLEKNATDN 350
Db 346 VKTM-KDILSKVNTYIDFNK---KYKTVNSEHISQFTLTDKIKAEVSDKELKK----- 396
QY 351 ISKLPFAPPSHEETDS-TKEAAKMEKYGSLKSDKSDNSPGGKTDPKPKTRAYL 409
Db 397 -----CEQSFNDNKLINETKNSIEKEY-----QNINTLKKVDE-----YI 432
QY 410 EAIRKNIEWLKHDKGKNEDYDLSKMDPFIKQADAYVEKGILDKEAEAIKRI 464
Db 433 KYVKSTKESITKFSK-----QTILKMLNQNIKTVKETSIDKSYIEKFEQI 480

RESULT 13
A43336
microtubule-vesicle linker CLIP-170 - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C:Accession: A43336
R:Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A:Title: CLIP-170 links endocytic vesicles to microtubules.
A:Reference number: A43336; MUID:92405160; PMID:1356075
A:Accession: A43336
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1392 <PE>
A:Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

Query Match 7.4%; Score 178.5; DB 2; Length 1392;
Best Local Similarity 22.4%; Pred. No. 0.42;
Matches 111; Conservative 79; Mismatches 182; Indels 123; Gaps 20;

QY 34 RELSAB-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNVSPVDNL-----NLKA 80
Db 734 RKASSEKSEMKLRQQL-EAAEQIKHLEIEKNAESSKASSITRELQGRELKLNTNLOEN 792
QY 81 ITE-----KEKIEKERQ-----SIRSPLD--NKL-----NVEDVDSTKN 113
Db 793 LSEVSQVKETLEKELQILKEKFAEASEAVSVQSMQETVKNLHQKEEQFNMSSDLEKL 852
QY 114 RKLIDDYDSTKSGLDHKFQDDPGLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVS 173
Db 853 RENLAD-----MEAKFREKDEREEQL-----IKAKEKLEND-----IA 885
QY 174 KLLNLGLITESQAHTLEDEV-----AEVLQKLISKEANNYE-----EDPNKPTSWTEN 221
Db 886 EIMKMGDNGNSQLTKMDELRLKERDVEELQKLTKANENASFLQKSIDMTVKAESQSQ 945
QY 222 QAGKIPEKVTMAAIQDGLAKGENDETVSNLTILTNGLERRTKTYSEDNFEELQYFPNFY 281
Db 946 EAAKKHBE-----EKKELERKLSL-----LEKQMET-SHNQCOELK----- 980
QY 282 ALLKSIDSEKAEKETLITIMKTLPIDFVMMVKYGTISPPEGVSYLENDEMIALOTKN 341
Db 981 ARYERATSETKHEEILQNLQKLTLDLTDKLG-----ARENSGLLQLEELRKRQAEKA 1036
QY 342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEAAKMEKYGSLKSDKSDNSPGGKTD 399
Db 1037 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNKQLNDELTLKENNL-KNVEELNKS 1095
QY 400 EPKGKTEAYLEAIRKNIEWLKH-----DKGNKEDYDLSKMDPFIKQADAYV 448
Db 1096 ELLTVENQKMEEFREKTEITLKQAAAKSQQLSALQENNVLAELGRSDRDEVTSHQKLEE 1155
QY 449 EKGILDKEAEAIKR 463
Db 1156 ERSVLNNQLLEMKKR 1170

```

```

RESULT 14
S22695
restin - human
C:Species: Homo sapiens (man)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999
C:Accession: S22695; S19853
R:Balbe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;
EMBO J. 11, 2103-2113, 1992
A:Title: Restin: a novel intermediate filament-associated protein highly expressed in the
A:Reference number: S22695; MUID:92289675; PMID:1600942
A:Accession: S22695
A:Molecule type: mRNA
A:Residues: 1-1427 <BI>
A:Cross-references: EMBL:X64838; NID:g35998; PIDN:CAA46050.1; PID:g35999
C:Keywords: cytoskeleton

```

```

Query Match 7.4%; Score 178.5; DB 2; Length 1427;
Best Local Similarity 22.4%; Pred. No. 0.43;
Matches 111; Conservative 79; Mismatches 182; Indels 123; Gaps 20;

QY 34 RELSAB-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNVSPVDNL-----NLKA 80
Db 769 RKASSEKSEMKLRQQL-EAAEQIKHLEIEKNAESSKASSITRELQGRELKLNTNLOEN 827
QY 81 ITE-----KEKIEKERQ-----SIRSPLD--NKL-----NVEDVDSTKN 113
Db 828 LSEVSQVKETLEKELQILKEKFAEASEAVSVQSMQETVKNLHQKEEQFNMSSDLEKL 887
QY 114 RKLIDDYDSTKSGLDHKFQDDPGLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVS 173
Db 888 RENLAD-----MEAKFREKDEREEQL-----IKAKEKLEND-----IA 920
QY 174 KLLNLGLITESQAHTLEDEV-----AEVLQKLISKEANNYE-----EDPNKPTSWTEN 221
Db 921 EIMKMGDNGNSQLTKMDELRLKERDVEELQKLTKANENASFLQKSIDMTVKAESQSQ 980
QY 222 QAGKIPEKVTMAAIQDGLAKGENDETVSNLTILTNGLERRTKTYSEDNFEELQYFPNFY 281
Db 981 EAAKKHBE-----EKKELERKLSL-----LEKQMET-SHNQCOELK----- 1015
QY 282 ALLKSIDSEKAEKETLITIMKTLPIDFVMMVKYGTISPPEGVSYLENDEMIALOTKN 341
Db 1016 ARYERATSETKHEEILQNLQKLTLDLTDKLG-----ARENSGLLQLEELRKRQADKA 1071
QY 342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEAAKMEKYGSLKSDKSDNSPGGKTD 399
Db 1072 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNKQLNDELTLKENNL-KNVEELNKS 1130
QY 400 EPKGKTEAYLEAIRKNIEWLKH-----DKGNKEDYDLSKMDPFIKQADAYV 448
Db 1131 ELLTVENQKMEEFREKTEITLKQAAAKSQQLSALQENNVLAELGRSDRDEVTSHQKLEE 1190
QY 449 EKGILDKEAEAIKR 463
Db 1191 ERSVLNNQLLEMKKR 1205

```

```

RESULT 15
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96796
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpsil, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

```

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1871 <STO>
A;Cross-references: GB:AE005173; NID:G6143906; PIDN:AAF0452.1; GSPDB:GN00141
C;Genetics:
A;Gene: F28O16.15
A;Map position: 1

```
Query Match      7.4%; Score 177; DB 2; Length 1871;
Best Local Similarity 20.7%; Pred. No. 0.71;
Matches 116; Conservative 85; Mismatches 212; Indels 148; Gaps 22;

QY 12 VLVLPITQAFPKP--GGSQDKSLHNRELSAERPLNEQIAEABEDKIKTYPPENKPGOSNY 69
Db 1052 VLVVEETYPKHTGGEHDNDHKEEQENVIKAAELNTEEDSPKVEEIEKQ----- 1105

QY 70 SFVDNLNLKAI TEKEIEKERSIRSPLOKLNVEDVDSTKN--RKLDDYDSTKSG 126
Db 1106 ---DHGELKRSVMVQAKRQTEEKD-KTRAMEKNETVERRKQTKDGLKREGEDPELGG 1161

QY 127 LDHKFQDDPDGLHQLDGTPLT--AEDIVHKTAARIYEENDRAVFD----KIVSKLLNLGL 180
Db 1162 --HERRGEEDRIEELVETEISDHKEKVKKDEDYILRSQDTGKVDLGERERRSKQRKIHK 1219

QY 181 ITESQAHTLEDEVAEVLQKLISKEAN-----NVEEDPNKPTSWTENQAGKIPKVT- 232
Db 1220 SVEDEIGQDEDEAEAAVSRNENGSSRKQVTEIESEK-----HKEQNKIPETSNPE 1274

QY 233 -----MAAIQDGLAKGEN-----DETVSNTLTLTNGLERRTKT 265
Db 1275 VNEEDEERVVEKETVEVAHQELEGKTENCKDDDGERRERCKQGTAEENMLRQRFKT 1334

QY 266 YSEDNFEELQYFPNFYALKSID--SEKEAKEKTLITIMKTLIDFVMMVKYGTI---- 319
Db 1335 KSDD-----GLVRKIQTKEEPEDEKKS---QESSSHVVKLVAEEDGSLRNL 1378

QY 320 ---SPEEGVSYLENLDEMIALQTKNLEKNATD--NISKLFPAPSEKSHETDST----- 369
Db 1379 EFSEKESTVSKMLKIDSEKKEEHKIRKPTTEERSNAPVIEKQGNKKNAAEENQDKIDRR 1438

QY 370 -KEEAQWKEKYGSLKD-----STKDDNSNFGGKT-- 399
Db 1439 GKNOEIKGQEPYGLVRNGEHDKI TEYHRGEEKGTAEENVSTKIQTKDELEKPKPSEIS 1498

QY 400 -----EPKGKTEAYLEAIRKNIEMLKKHDKKGNKEDYDLS--KMRDFIN 441
Db 1499 ENHNIHEFMDSSQSDIEEKGSQDAEKYAKQNKIQEVWDEDR--KEEYHISERVNEMA 1556

QY 442 K---QADAYVEKGILDKERAE 459
Db 1557 KRILQVESKANDGSSKKNETE 1577
```

Search completed: January 2, 2004, 12:36:54
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 18 Seconds
(without alignments)
1222.694 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTLVLVLPIQAF.....EKGILDKBBAIKRIYSSL 468

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2395	99.8	468	1	SG3_HUMAN	Q8Wxd2 homo sapien
2	2102.5	87.6	471	1	SG3_MOUSE	P47867 mus musculus
3	2096.5	87.4	471	1	SG3_RAT	P47868 rattus norv
4	191	8.0	2464	1	MAPB_MOUSE	P14873 mus musculus
5	186.5	7.8	2116	1	MY52_DICDI	P08799 dictyosteli
6	183.5	7.6	1790	1	USO1_YEAST	P25386 saccharomyc
7	180.5	7.5	2459	1	MAPB_RAT	P15205 rattus norv
8	179	7.5	1005	1	RA50_METJA	Q58718 methanococc
9	178.5	7.4	1427	1	REST_HUMAN	P30622 homo sapien
10	178	7.4	882	1	RA50_PYREFU	P58301 pyrococcus
11	172.5	7.2	1164	1	BAG_STRAG	P27951 streptococc
12	169	7.0	700	1	TRDN_CANPA	P82179 canis famil
13	167.5	7.0	957	1	KP5C_HUMAN	O60282 homo sapien
14	166.5	6.9	1875	1	MU1P_YEAST	Q2455 saccharomyc
15	166	6.9	2663	1	CENE_HUMAN	Q02224 homo sapien
16	166	6.9	5038	1	PCLO_MOUSE	Q9gyx7 mus musculus
17	163.5	6.8	956	1	KP5C_MOUSE	P280739 mus musculus
18	163	6.8	1251	1	RPB2_PLAYVB	O00799 plasmodium
19	161.5	6.7	1360	1	CING_XENLA	Q9ptd7 xenopus lae
20	161.5	6.7	1969	1	MYSA_CABEL	P22844 caenorhabdi
21	161	6.7	852	1	RA50_THEMA	Q9x1x1 thermotoga
22	159.5	6.6	997	1	CPH1_RAT	Q03410 rattus norv
23	159.5	6.6	1978	1	MYHB_CHICK	P10587 gallus gall
24	159	6.6	1208	1	PCP1_SCHPO	Q92351 schizosacch
25	158.5	6.6	1526	1	MY52_SCHPO	Q9usi6 schizosacch
26	158	6.6	1972	1	MYHB_RABIT	P35748 cryptolagus
27	157.5	6.6	1972	1	MYHB_HUMAN	P35749 homo sapien
28	157.5	6.6	2230	1	GOG4_HUMAN	Q13439 homo sapien
29	157	6.5	727	1	MPF1_ARATH	Q91w85 arabidopsis
30	157	6.5	944	1	NUF1_YEAST	P32380 saccharomyc
31	157	6.5	2017	1	MY5N_DROME	Q93323 drosophila
32	156.5	6.5	1087	1	AKA9_RABIT	Q28628 cryptolagus
33	156.5	6.5	2468	1	MAPB_HUMAN	P46821 homo sapien

34	156	6.5	705	1	TRDN_RABIT	Q28820 oryctolagus
35	155.5	6.5	1130	1	YLI1_CABEL	Q11102 caenorhabdi
36	155.5	6.5	5085	1	PCLO_RAT	Q9jks6 rattus norv
37	154.5	6.4	1539	1	Y373_HUMAN	O15078 homo sapien
38	154	6.4	1433	1	REST_CHICK	O42184 gallus gall
39	154	6.4	1957	1	SPOF_SCHPO	O10411 schizosacch
40	154	6.4	1962	1	MYSA_DROME	P05661 drosophila
41	153.5	6.4	1976	1	MYHA_RAT	Q9j1c0 rattus norv
42	153.5	6.4	2104	1	MY53_SCHPO	O14157 schizosacch
43	152.5	6.4	2022	1	ANTI_ONCVO	P21249 onchocerca
44	152.5	6.4	8545	1	ANCI_CABEL	Q9n4m4 caenorhabdi
45	152	6.3	667	1	CYLI_BOVIN	P35662 bos taurus

ALIGNMENTS

RESULT 1

ID	SG3_HUMAN	STANDARD;	PRT;	468 AA.
AC	Q8Wxd2; Q96C83; Q96GE8; Q9Y6G7;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Secretogranin III precursor (SgIII).			
GN	SG3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreatic;			
RA	Song H., Peng Y., Huang Q., Dai M., Mao Y., Zhang Q., Mao M., Fu G.,			
RA	Luo M., Chen J., Hu R.;			
RT	"Human secretogranin III mRNA, complete cds";			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.			
RX	MEDLINE=22093564; PubMed=12098761;			
RA	Rong Y.P., Liu F., Zeng L.C., Ma W.J., Wei D.Z., Han Z.G.;			
RT	"Cloning and characterization of a novel human secretory protein:			
RT	secretogranin III.";			
RN	Acta Biochim. Biophys. Sin. 34:411-417(2002).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain, and Lung;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahney J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,			
RA	Schneerch A., Schein J.F., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
CC	-!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine			
CC	secretory granules.			
CC	-!- TISSUE SPECIFICITY: Expressed in brain, heart, kidney, liver and			
CC	skeletal muscle.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; AF078851; AAD44483.1; -
 DR EMBL; AF453583; AAL67431.1; -
 DR EMBL; BC014539; AAH14539.1; -
 DR EMBL; BC009511; AAH09511.1; AUT_INIT.
 DR Genew; HGNC:13707; SCG3.
 KW Signal; Cleavage on pair of basic residues; Polymorphism.
 FT SIGNAL 1 19
 FT CHAIN 20 468
 FT VARIANT 125 125
 FT CONFLICT 79 79
 FT CONFLICT 167 167
 FT CONFLICT 272 274
 FT SEQUENCE 468 AA; 52977 MW; 633A87692A783808 CRC64;
 SQ

Query Match 99.8%; Score 2395; DB 1; Length 468;
 Best Local Similarity 99.8%; Pred. No. 6e-108;
 Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGFLGTGWLVLVLPQAPKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKTYP 60
 DB 1 MGFLGTGWLVLVLPQAPKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKTYP 60
 QY 61 ENKPGQSNYSFVDNLNLLKAIETKEKIERQSISSPLDNKLNVEDVDSTKRNKLIDDY 120
 DB 61 ENKPGQSNYSFVDNLNLLKAIETKEKIERQSISSPLDNKLNVEDVDSTKRNKLIDDY 120
 QY 121 DSTKSGLDHKKFODPDGHLQDGLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
 DB 121 DSTKSGLDHKKFODPDGHLQDGLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
 QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTMPAAIQDGL 240
 DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTMPAAIQDGL 240
 QY 241 AKGENDETVENTLTNLGLERTKTYSEDNFEELQYFPNFPYALLKSIDSEKAKEKETLI 300
 DB 241 AKGENDETVENTLTNLGLERTKTYSEDNFEELQYFPNFPYALLKSIDSEKAKEKETLI 300
 QY 301 TIMKTLIDFVMMVKYGTISPGEVSVLENLDEMIALQTKNKLKRNATDNISKLPAPSE 360
 DB 301 TIMKTLIDFVMMVKYGTISPGEVSVLENLDEMIALQTKNKLKRNATDNISKLPAPSE 360
 QY 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDSDNSNPGGKTDEPKGTEAYLEAIRKNIEWLK 420
 DB 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDSDNSNPGGKTDEPKGTEAYLEAIRKNIEWLK 420
 QY 421 KHKDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
 DB 421 KHKDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 2

SG3_MOUSE
 ID SG3_MOUSE STANDARD; PRT; 471 AA.
 AC P47867;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Secretogranin III precursor (SgIII).
 GN SCG3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RN SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=95001263; PubMed=7917832;
 RA Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
 RA Sutcliffe J.G.;
 RT "Primary structure of mouse secretogranin III and its absence from
 RT deficient mice";
 RL J. Mol. Neurosci. 4:225-233(1993).
 CC -|- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
 CC secretory granules.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; U02982; AAA56636.1; -
 DR MGD; MGI:103032; SCG3.
 KW Signal; Cleavage on pair of basic residues.
 FT SIGNAL 1 22
 FT CHAIN 23 471
 FT SEQUENCE 471 AA; 53326 MW; FCIE9C381AFCAS64 CRC64;

Query Match 87.6%; Score 2102.5; DB 1; Length 471;
 Best Local Similarity 88.3%; Pred. No. 5.8e-94;
 Matches 416; Conservative 17; Mismatches 35; Indels 3; Gaps 1;

QY 1 MGFLGTGWLVLVLPQAPKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKT 57
 DB 1 MGFLGTGWLVLVLPQAPKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKKA 60
 QY 58 YPENKPGQSNYSFVDNLNLLKAIETKEKIERQSISSPLDNKLNVEDVDSTKRNKLI 117
 DB 61 PPSEKPSSESNSYSSVDNLNLLRAITEKETVEKERQSISSPPFDQNLVEDDADSTKRNKLI 120
 QY 118 DDDYSTKSGLDHKKFODPDGHLQDGLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNL 177
 DB 121 DEYDSTKSGLDHKKFODPDGHLQDGLDGTPLTAEDIVHKIAARIYEENDRGVDFKIVSKLNL 180
 QY 178 LGLITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTMPAAIQ 237
 DB 181 LGLITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTMPAAIQ 240
 QY 238 DGLAKGENDETVENTLTNLGLERTKTYSEDNFEELQYFPNFPYALLKSIDSEKAKEKE 297
 DB 241 DGFTNRENDDETVENTLTNLGLERTKTYSEDNFEELQYFPNFPYALLKSIDSEKAKEKE 300
 QY 298 TLITIMKTLIDFVMMVKYGTISPGEVSVLENLDEMIALQTKNKLKRNATDNISKLPAP 357
 DB 301 TLITIMKTLIDFVMMVKYGTISPGEVSVLENLDEMIALQTKNKLKRNATDNISKLPAP 360
 QY 358 PSEKSHETDSTKEEAAKMEKEYGSLKSDTKDSDNSNPGGKTDEPKGTEAYLEAIRKNIE 417
 DB 361 PPEKSQEBETDSTKEEAAKMEKEYGSLKSDTKDSDNSNPGGKTDEATGTEAYLEAIRKNIE 420
 QY 418 WLKHKDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
 DB 421 WLKHKDKGNKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 471

RESULT 3

SG3_RAT
 ID SG3_RAT STANDARD; PRT; 471 AA.
 AC P47868;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Secretogranin III precursor (SgIII) (151075).
 GN SCG3.

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95001263; PubMed=7917832;
 RA Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
 RA Sutcliffe J.G.;
 RT "Primary structure of mouse secretogranin III and its absence from
 RT deficient mice.";
 RL J. Mol. Neurosci. 4:225-233(1993).
 RN [2]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=90376160; PubMed=2204688;
 RA Ottiger H.-P., Battenberg E.F., Tsou A.-P., Bloom F.E.,
 RA Sutcliffe J.G.;
 RT "18075: a brain- and pituitary-specific mRNA that encodes a novel
 RT chromogranin / secretogranin-like component of intracellular
 RT vesicles.";
 RL J. Neurosci. 10:3135-3147(1990).
 CC -1- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
 CC secretory granules.
 CC -1- TISSUE SPECIFICITY: BRAIN- AND PITUITARY-SPECIFIC.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U02983; AAA56637.1; -;
 KW SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 471 SECRETOGRANIN III.
 SQ SEQUENCE 471 AA; 53183 MW; 977B3F8885F33223 CRC64;
 Query Match 87.4%; Score 2096.5; DB 1; Length 471;
 Best Local Similarity 87.9%; Pred. No. 1.1e-93;
 Matches 414; Conservative 18; Mismatches 36; Indels 3; Gaps 1;
 QY 1 MGFLGTGWILVLVL---PIQAFPKPGSQKSLHNRSLAERPLNEQIAEADKIKKT 57
 DB 1 MGFLWTGSWILVLVNSGPIQAFPKPGSQKSLHNRSLAERPLNEQIAEADKIKKT 60
 QY 58 YPPENKPGQSNYSFVDNLNLKAIKTEKEIKERQSRSSPLDNKLVNEDVDSTKNRKL 117
 DB 61 YPSESKEPSESFNFSVDNLNLKAIKTEKETVEKAKQSRSSPFDNRLNVDDADSTKNRKL 120
 QY 118 DDYSTKSLGDKHKKQDDPDGLHOLDGTLTAEDIVHKIARIYEENDRAVDFKIVSKLLN 177
 DB 121 DEYDSTKSLGDKRVQDDPDGLHOLDGTLTAEDIVHKIARIYEENDRGVDFKIVSKLLN 180
 QY 178 LGLITESQAHTLEDAVAVLOKLSKEANNYEEDPNKPTSWENQAGKIPKVTPEMAIQ 237
 DB 181 LGLITESQAHTLEDAVAVLOKLSKEANNYEEDPEKPTSTENQDGKIPKVTPEVATQ 240
 QY 238 DGLAKGENDETSVNTLTNLNGLRRTKTSYEDNFEELQYFFNFYALLKSIDSEKEAKEKE 297
 DB 241 DGFTNRENDTVSNTLTNLNGLRRTNPHRDDDFEELQYFFNFYALLTSIDSEKEAKEKE 300
 QY 298 TLTIIMKTLIDFVKMWKYGTISPEGVSYLENDEMIALQTKNLEKNTADNTSKLPPA 357
 DB 301 TLTIIMKTLIDFVKMWKYGTISPEGVSYLENDETIALQTKNLEKNTDTSKLPFA 360
 QY 358 PSEKSHETDSTKEAAKWEKEYGSLKSDTDDNSNPGKTDPEKGTKEAYLEAIRKNIE 417
 DB 361 PPEKSHETDSTKEAAKWEKEYGSLKSDTDDNSNPGKTDPEKGTKEAYLEAIRKNIE 420
 QY 418 WLKXHDKKGKEDYDLSKQDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

Db 421 WLKXHNKGNKEDYDLSKQDFINKQADAYVEKGILDKKEAEAIKRIYSSL 471
 RESULT 4
 MAPB_MOUSE
 ID MAPB_MOUSE STANDARD; PRT; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Last created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
 DE [Contains: MAP1 light chain LC1].
 GN MAP1B OR MTAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DOMAIN.
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RA Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein
 RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
 RT and tau.";
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
 CC OF MAP1B.
 CC -1- SIMILARITY: TO MAP1A.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X51396; CAA35761.1; -;
 DR PIR; S07549; QRMSP1.
 DR MGD; MGI:1306778; Mtap1b.
 DR GO; GO:0016358; P:dendrite morphogenesis; IMP.
 DR GO; GO:0001578; P:microtubule bundling; IMP.
 DR InterPro; IPR00102; MAP1B_neuraxin.
 DR Pfam; PF00414; MAP1B_neuraxin; 10.
 DR PROSITE; PS00230; MAP1B_NEURAXIN; 7.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
 FT REPEAT 1874 1890 MAP1B 1.
 FT REPEAT 1891 1907 MAP1B 2.
 FT REPEAT 1908 1924 MAP1B 3.
 FT REPEAT 1925 1941 MAP1B 4.
 FT REPEAT 1942 1958 MAP1B 5.
 FT REPEAT 1959 1975 MAP1B 6.
 FT REPEAT 1993 2009 MAP1B 7.
 FT REPEAT 2010 2026 MAP1B 8.
 FT REPEAT 2027 2043 MAP1B 9.
 FT REPEAT 2044 2060 MAP1B 10.
 FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY


```

RESULT 6
USOI_YEAST
ID USOI_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOI.
GN USOI OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VDP/USOI/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54378; CAA38253.1; --
DR EMBL; L03188; AAB00143.1; --
DR EMBL; U53668; AAB66659.1; --
DR SGD; S0002216; USOI.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol_p115_C.
DR InterPro; IPR006953; Usol_p115_head.
DR Pfam; PF04871; Usol_p115_C; 1.
DR Pfam; PF04869; Usol_p115_head; 1.
DR Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
DR DOMAIN 1 724
DR DOMAIN 725 1790
DR DOMAIN 465 487
DR DOMAIN 991 1790
DR DOMAIN 1172 1786
DR CHARGED (HYPER-HYDROPHILIC).
DR DISPENSABLE FOR THE PROTEIN FUNCTION.
DR ASP/GLU-RICH (ACIDIC).
DR G -> E (IN REF. 2).
DR E -> K (IN REF. 2).
DR V -> I (IN REF. 2).
DR I -> V (IN REF. 2).
DR N -> S (IN REF. 2).
DR G -> S (IN REF. 2).
DR I -> V (IN REF. 2).
DR I -> S (IN REF. 2).
DR D -> DEEDDEE (IN REF. 2).

```

```

SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B21659FD4818 CRC64;
Query Match 7.6%; Score 183.5; DB 1; Length 1790;
Best Local Similarity 20.3%; Pred. No. 0.15;
Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;
QY 29 KSLHN--RELSAERPLNEQIAAEEDKIKKTYPPENKPGQSNVFDVNLNLLKAI---T 82
DB 965 KSLANNYKMQAE---NESLIKAVEE-----SKNESSIQLSNLQNKQSDMSQ 1008
QY 83 EKEKIEKERSIRSSPLDNKLVNVEDVSTKNRKLI-----DDYSTKSGLDHKKFD--- 133
DB 1009 EKENFOIERGSEIKNIEQLKTTISDLQTKKEIISKSDSSKDEYSQISLLKLELTATT 1068
QY 134 -DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH----- 187
DB 1069 ANDENVNKISIELTKTRELEAEALAA--YKNLKNLELTLETKSEKALKEVNEEHLKEEK 1126
QY 188 -TLEDEVAEVLQKLISKAN-----NYEED--PNKPTSWTEQAGKIPEK 229
DB 1127 IQLEKEATETKQOLNSLRANLESLEKEHEDLAAQLKYYEQIANKERQYNE-EISQLNDE 1185
QY 230 VTPMAIQDGLAKGENDETVSNTLTITNGLRRT-----KTYSEDN-- 270
DB 1186 ITSTQOENESIKK-KNDELEGEVKAAMKSTSEQSNLKKSEIDALNLQIKELKKKNETNEA 1244
QY 271 -----FEELOYFPNF-----YALLKSIDSEKAK 294
DB 1245 SLLESIKSVESVETVKKLQDCNFKPEKSEVSELDKPKASEDKNSKYLELQK-ESEKINE 1303
QY 295 EKETLITIMKTLLDFVGMVK-----YALLKSIDSEKAK 315
DB 1304 ELDAKTELKIQLEKITNLSKAKESSELSLKKTSSEERKNAEQLKLNQIKNQ 1363
QY 316 -----YGTISPE--EGVSYLENDEMIALQTKNLEKVNATONISKLPAPSEK 361
DB 1364 AFERKLLNEGSTITQISEKINTLE--DELIRLQNEINELKAKEIDNTRSELEKVSLS 1421
QY 362 SHEETDSTKEEAAKKEVGSIKDS-TKDDNSNPGKTDPE-----KGKTEAYLEAIRK 414
DB 1422 NDLLEEEKQNTIKSQDLEILSYKDKITRNDEKLJSIERDKRDLSEKQELRAAQESKAK 1481
QY 415 NIEMLEKHKHKKGNEDYDLKMRDPIFKQADAYVEKGLD-KEEAFAIKR 463
DB 1482 VEEGLKLEESSEKAELEKSKEMM-KKLESTTIESNETLKSSNETIRK 1530
RESULT 7
MAPB_RAT
ID MAPB_RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
OS Rattus norvegicus (Rat).
GN MAP1B.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A.; DOMAIN, AND INDUCTION.
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;

```


RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
 RT "Identification of two distinct microtubule binding domains on
 RL recombinant rat MAP 1B.";
 RN Eur. J. Cell Biol. 57:66-74 (1992).
 [3]
 RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
 RC TISSUE-Spinal cord;
 RX MEDLINE=90059871; PubMed=2555150;
 RA Rientz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
 RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
 RT "Neuraxin, a novel putative structural protein of the rat central
 RT nervous system that is immunologically related to microtubule-
 RT associated protein 5.";
 RL EMBO J. 8:2879-2888 (1989).
 [4]
 RN DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
 RX MEDLINE=97405699; PubMed=9260743;
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
 RT in rat CNS and PNS during development.";
 RL J. Neurosci. Res. 49:319-332 (1997).
 CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
 CC that accompany neurite extension. Possibly MAP1B binds to at least
 CC two tubulin subunits in the polymer, and this bridging of subunits
 CC might be involved in nucleating microtubule polymerization and in
 CC stabilizing microtubules.
 CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
 CC with MAP1A and MAP1B proteins.
 CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
 CC heart or muscle.
 CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
 CC nerve levels are high early in development but decrease during
 CC postnatal development and are low in adults. In dorsal root
 CC ganglia levels remain high throughout development.
 CC -!- INDUCTION: By nerve growth factor.
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (By similarity).
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: TO MAP1A.
 CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
 CC 2459) was originally described as neuraxin in Ref.3.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; U52950; AAB17068.1; -;
 DR EMBL; X60370; CAC16162.1; -;
 DR EMBL; X16623; CAA34620.1; ALT_SEQ.
 DR FIR; A56577; A56577.
 DR InterPro; IPR000102; MAP1B neuraxin.
 DR Pfam; PF00414; MAP1B neuraxin; 10.
 DR PROSITE; PS00230; MAP1B NEURAXIN; 8.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2459
 FT MAP1 LIGHT CHAIN LC1.
 FT REPEAT 1869 1885 MAP1B 1.
 FT REPEAT 1886 1902 MAP1B 2.
 FT REPEAT 1903 1919 MAP1B 3.
 FT REPEAT 1920 1936 MAP1B 4.
 FT REPEAT 1937 1953 MAP1B 5.
 FT REPEAT 1954 1970 MAP1B 6.
 FT REPEAT 1988 2004 MAP1B 7.

FT REPEAT 2005 2021 MAP1B 8.
 FT REPEAT 2022 2038 MAP1B 9.
 FT REPEAT 2039 2055 MAP1B 10.
 FT DOMAIN 559 1035 GLU-RICH.
 FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 FT KKEE AND KKEI/V REPEATS).
 FT LYS-RICH.
 FT DOMAIN 2224 2312 M -> V (IN REF. 1).
 FT CONFLICT 127 127 T -> S (IN REF. 1).
 FT CONFLICT 140 140 R -> K (IN REF. 3).
 FT CONFLICT 2112 2112 L -> I (IN REF. 3).
 FT CONFLICT 2169 2169 L -> I (IN REF. 3).
 SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DDEB8BA2 CRC64;
 Query Match 7.5%; Score 180.5; DB 1; Length 2459;
 Best Local Similarity 22.7%; Pred. No. 0.31;
 Matches 120; Conservative 77; Mismatches 200; Indels 131; Gaps 22;
 Qy 22 KPGSQDQLNRLNRLSARPLNEQIAEAEDKIK-KYPPENKPGSQSYFVNLNLKA 80
 Db 549 KPLSSKSVRKESKEAPEATKASQVETPKVESKEKIVVKDKPKG-----VESKPS 600
 Qy 81 ITEKEKTEKERSIRSSPLDNKLNVDDVSTKNKLIDDVDTKSLDGHKFQDDP--DGL 138
 Db 601 VTEKEVSKESQSPVKAFAEAKAATESKPKVKDKVVKKEIKTP--EEKKEEPKKEVA 658
 Qy 139 HQLDGTPLTADIVHKIAARIYEENDRAVFDKIVSKLLNGLLITESTQAHT-LEDEVAEVL 197
 Db 659 KKEDKTPLKXDEKPKK-----EAKKEIKKEIKKEKELKKEVKETPLKDAKKEV- 710
 Qy 198 QKLISKANVYEDPNK-----PTSTENQAG-----KPEKVTPTMAA- 235
 Db 711 KKDEKKEVKEEKEPKKEIKKIDIKKSTPLSDTKKPAALKPKVAKKEEPTKKEPIAAG 770
 Qy 236 -IOD-GLAK-----GENDTVSNTL-----TLTNGLERRTK 264
 Db 771 KLKDKGVKVIKKEGKTEAAATAGTAATAAAGVAAAGPAKLEAERSLMSSEPDLT 830
 Qy 265 TYSEDNFEELOYPFNFYALLKSIDSEKAKE-----KETLIITIMKTLIDFVKMMV 314
 Db 831 DFELKAEIDVAKDIKPQLLEIEDEKLEKTEPEGAUVIQTKEVS-----877
 Qy 315 KYGTISPEEGSVYLENDLMALQTKNL---EKNATDNISKL-----PPAPEKSHHE 365
 Db 878 KGSAESDEGITTTEGESE--CEQTPPEEPVEKQGVDDIEKFEDEGAFESSEAGDYE 935
 Qy 366 TDSKTEAAKWEKEYGSLKSDTKDDNSNPGKTDPEKGTKEA--YLEAIRKNI-----416
 Db 936 EKAETEAEPEEDG-----EDNVSGSASKHSPTDEETIAKAEADVHIKKEKRESVAGSDRA 992
 Qy 417 -----EWLKKHKKGNKEDYDLKMRDFINKQADAYVEKGLDKKEAE 459
 Db 993 EEDMDALEKGEAEQSEEGESEE-----DKAEDAREEDHEPDKTEAE 1035
 RESULT 8
 RA50 METJJA
 ID RA50 METJJA STANDARD; PRT; 1005 AA.
 AC Q58718;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA double-strand break repair rad50 ATPase.
 GN RAD50 OR MJ1322.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RC MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Smith M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67572; AAB99331.1; --
CC PIR; A64465; A64465.
CC TIGR; MJ1322; --
CC HAMAP; MF_00449; --
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR003405; SMC C.
CC InterPro; IPR003395; SMC N.
CC Pfam; PF04423; Rad50 zn hook; 1.
CC Pfam; PF02483; SMC C7; 1.
CC Pfam; PF02463; SMC N; 1.
CC ProDom; PD000006; ABC transporter; 1.
CC SMART; SM00382; AAA; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP BIND 32 39
CC FT DOMAIN 158 849 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E798F3 CRC64;

Query Match 7.5%; Score 179; DB 1; Length 1005;
Best Local Similarity 22.3%; Pred. No. 0.12;
Matches 116; Conservative 80; Mismatches 164; Indels 160; Gaps 27;

QY 52 DKIKKTPPENKPGQSNYSFVNLNLKA-ITKEKIEKQRQIRSPDLNKLN-VEVDV 109
DB 159 DEFECY---QKMGIEVKEKRLERIEGELNYKENYKE-----LNKMSQLEE-- 205
QY 110 STNKRKLIDDYDTKSLGDKHFQDDPGLHQLDGTPLTAEDIVHKIAARIYENDRAVDF 169
DB 206 --KNKLMEINDKLNK-IKKEFEDIEKLFNENWKLLEYKFNKL-----BERKALEL 257
QY 170 KIVS-KLLNGLTETSAQH-----TLEDEVAVLQKLIKEANNYEEDPNK 214
DB 258 KNGELKILEYDLNTVVEARETLNRHKDEVEKYKSLVDEIRKIESRL--RELKSHVEDYLK 315
QY 215 PTGWTENQAGKIPKVTPTMAAIOGKAGNDETVNTLTITNGLERRTKTYSNDFEEL 274
DB 316 LTQLEIKGDI-EK-----LKEFINKKYRDDIDNLDLTLNKI-----KDEIERV 360
QY 275 QYFNFVALLKSIDSEKAEK-----ETLITIMKTLIDFVQVMKYGTISPE 322
DB 361 ETIKDLLEELKNLEIEKIEKICEKEVEYKYLEEKAVEYNKLTLEYITL--- 417
QY 323 EGVSYLENDEMTALQTKNLEKATD---NISKLPAPS-----EKSHEETSTK-- 370
DB 418 -----LOEKKSIEKINDLETRINKLEETKNIDIESIENSLKEIEKKV 463
QY 371 -----EEAAKMEKEYGLKSDTK-----DDNSNPGK-----TDEPK-----GKT 405

Db 464 LENLQKEKIELANKLGEINSEIKRLKKILDKEVEGKPLCKTPTIDENKQMLINQHK 523
QY 406 E---AYLE-----AIRKNIEMLKKH-----DKGKNKEDYDL----- 433
DB 524 QLNKKYTELEENKKTRETEKIEKLEKIDKEENLKTLYLEKQSQIELEELKKNV 583
QY 434 SKMRDPINKQADAYVEKG-----ILD--KEEAIAIKRIYS 466
DB 584 KEQLDEINKKISNYVINGKPVDEILEDIKSQLNKPKNFYN 623

RESULT 9
REST_HUMAN
ID REST_HUMAN STANDARD; PRT; 1427 AA.
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
DE Sternberg intermediate filament associated protein).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Peripheral blood monocytes;
RX MEDLINE=92289675; PubMed=1600942;
RA Bilbe G., Delabie J., Brueggen J., Richener H., Asseberg F.A.M.,
RA Cerletti N., Sorg C., Odink K., Tarsay L., Wiesendanger W.,
RA de Wolf-Peters C., Shipman R.;
RA "Restin: a novel intermediate filament-associated protein highly
RT expressed in the Reed-Sternberg cells of Hodgkin's disease";
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92405160; PubMed=1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RA "CLIP-170 links endocytic vesicles to microtubules";
RL Cell 70:887-900(1992).
CC -!- FUNCTION: SEEMS TO BE AN INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES
CC -!- SURCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P30622-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P30622-2; Sequence=VSP_000765;
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64838; CAA46050.1; --
CC EMBL; M97501; AAA35693.1; --
CC PIR; S22695; S22695.
CC Genew; HGNC:10461; RSN.
CC MIM; 179838; --
CC GO; GO:0005768; C:endosome; TAS.
CC GO; GO:0005882; C:intermediate filament; TAS.
CC GO; GO:0015630; C:microtubule cytoskeleton; TAS.
CC GO; GO:0008017; F:microtubule binding activity; TAS.

```
DR GO: 0006899; P: non-selective vesicle transport; TAS.
DR InterPro: IPR000938; CAP-Gly.
DR Pfam: PF01302; CAP_GLY_2.
DR SMART: SM00343; Znf_C2HC; 1.
DR PROSITE: PS00845; CAP_GLY_1; 2.
DR PROSITE: PS0245; CAP_GLY_2; 2.
KW Cytoskeleton, Microtubules; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 78 120 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT DOMAIN 232 274 CAP-GLY 2.
FT DOMAIN 304 331 SER-RICH.
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
FT DOMAIN 1408 1421 CCHC-BOX.
FT VARSPIC 457 491 Missing (in isoform Short).
FT CONFLICT 1069 1069 /FTD=VSP_000765.
FT SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match
Best Local Similarity 22.4%; DB 1; Length 1427;
Matches 111; Conservative 79; Mismatches 182; Indels 123; Gaps 20;

QY 34 RELSAG-----RPLNEQIABAEEDKIKKTPPENKPGOSNYFVDNL-----NLKA 80
DB 769 RASSEGKSEKWKLRQQL-EAAEQIHLKLEKKAESSKASSITRELQGRLEKLTNIQEN 827
QY 81 ITE-----KEKIEKQ-----SIRSPLD--NKL-----NVEDVDSTKN 113
DB 828 LSEVSQVKETLEKELQILKFAEABEAVSQSMQETVKNLKHQKEQFNWLSDDLKL 887
QY 114 RKLDDYDSTKGLDHKFDQDPDGLHLDGTPLTAEDIVHKIARIYEENDRAVDFKIVS 173
DB 888 RENLAD-----MEAFKREDEREQ-----IKAKEKLEND-----IA 920
QY 174 KLLNLGLITESQAHTLEDEV-----ARVLQKLISKEANNYE-----EDPNKPTSWTEN 221
DB 921 EIMKSGNSQLTKWMDLKLKRDVEELQKLTUKANENASFLQKSIEDTIVKAEOSQ 980
QY 222 QAGKIPKVTMAAIQDGLAGXENDETSTNLTITNGLERTKYTSBDFEELQYFPNFY 281
DB 981 EAAKHED-----EKLEKRLSD-----LEKQWET-SHNQCELEK----- 1015
QY 282 ALLKSIDSEKAKETLITIMKTLIDFVKRMVKYGISPEGVSYLENLDEMIALQTKN 341
DB 1016 ARYERATSETKTKHEEILQNLQTLTLDTEKLG-----ARENSGLLQLEELKQADKA 1071
QY 342 KLEKNATDNISKLPAPSEKSH--EETDSTKEAAKMEKYGSLKSTKDDNSNPGKTD 399
DB 1072 KAAQTAEADAMQIMEQMTKEKTETTLASUEDTKTNKQLQNLDTLKNNL-KNVEELNKS 1130
QY 400 EPKGKTEAYLEAIRKNIEWLKKH-----DKGKNEDYDLSKWRDPINKQADAYV 448
DB 1131 ELLTVENQKMEFEKFEITLQAAAKSQQLSALQENVKLAELGSRDEVTSHQKLEE 1190
QY 449 EKGILDKEEAFAIKR 463
DB 1191 ERSVLNNQLEMKKR 1205

RESULT 10
RA50_PVRFU STANDARD; PRT; 882 AA.
AC P58301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PF1167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
```

```

[1] SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
RA Carney J.P.;
RT "Mrell and Rad50 from Pyrococcus furiosus: cloning and biochemical
RT characterization reveal an evolutionarily conserved multiprotein
RT machine.";
RT J. Bacteriol. 182:6036-6041(2000).
[2]
RC SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RC Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RA "The complete sequence of the Pyrococcus furiosus genome.";
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
[3]
RC X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20348838; PubMed=10892749;
RA Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
RA Carney J.P., Tainer J.A.;
RT "Structural biology of Rad50 ATPase: ATP-driven conformational
RT control in DNA double-strand break repair and the ABC-ATPase
RT superfamily.";
RT Cell 101:789-800(2000).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site.
CC -1- SUBUNIT: Forms a complex with mrell.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE010225; AL081291.1;
CC PDB; 1F2T; 20-SEP-00.
CC PDB; 1F2U; 02-AUG-00.
CC PDB; 1I18; 30-MAY-01.
CC PDB; 1L8D; 28-AUG-02.
CC HAMAP; MF_00449; -.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR003395; SMC N.
CC Pfam; PF04423; Rad50 zn hook; 1.
CC Pfam; PF02463; SMC N; 1.
CC ProDom; PD000006; ABC transporter; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
KW Complete proteome.
FT NP BIND 30 37 ATP.
FT DOMAIN 148 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 882 AA; 103839 MW; 3ADCBD250382A99E CRC64;

Query Match
Best Local Similarity 22.1%; DB 1; Length 882;
Matches 112; Conservative 90; Mismatches 172; Indels 132; Gaps 21;

QY 39 ERPLNEQIAEAE--DKIK---KTYPPENKPGOSNYFVDNLNLKA-----ITEKEKIE 88
DB 260 KRTLEBRKNTVEYLEKLEKEKELEBEQVKETISIKKQVDAYLALKEFRNEYLDKKYKE 319
QY 89 KERQSISSPLDNKLVNVEDVSTKNKRLIDDYDSTKGLDHKFDQDPDGLHLDGTPLTA 148
DB 320 KE--LTFVEELINEIQ-----KRIEELNEKESEKEKLENE-----K 353
QY 149 EDIVHKIARIYEENDRAVDFKIVSKLLNLGLITESQAHTLEDEVAVLEKLSKEANNY 208
DB 149 EDIVHKIARIYEENDRAVDFKIVSKLLNLGLITESQAHTLEDEVAVLEKLSKEANNY 208
```

Db 354 KEILNKLAAIL---EKDQHYEEIKAKKENLRQLKEKLGDKSPEDIKKLEEBETKK-TTI 409
 QY 209 BEDPNKPTSWTENQAKIPEKTPMAAQDGL--AKGE-----NDE----- 247
 Db 410 EBERNEIT---ORIGELKNKIGDGLKTAIEELKKAAGKPCVCGRELTDHEREELLSKYHL 465
 QY 248 -----TVSNLTILTLNGLERTKTYSDENPELOVFPNFYALLKSID-----E 290
 Db 466 DLNNSKNVLAKLIDRKSELERELRIDME-IKLTPLTVAEQIRSIIEELNVNLEKIE 524
 QY 291 KEAKEKEFLITIMTKLIDFVQKMKYKGTISPBEVSVYLENDEMIA--LQTKNKLKNAT 348
 Db 525 KNATYKLEELRLTEGRIGLA-----EDLKLAPLEKKAALIHKKQLEKELK 576
 QY 349 DNISKLFPAPSKSHEETDSTKEAAKMEKVGSLKDTKD-----DNSNPGSKTD 399
 Db 577 ELNTKL-ESFGKSVEDLSDKLELEEIYKRYLTLLNSKKELEITQREIAKAKETLEMSF 635
 QY 400 EPKGTEAYLEAIRKNIENLWKKHDKKGNEDY-----DLSKMR 437
 Db 636 BELAEVEADIERIEKLSQLK--QKYNEEYKXKREKELEKELARLEAQKLEKRR 692
 QY 438 DFINKQADAYVEKGLDKEEAHAKR 463
 Db 693 DTIK-----STLEKLAKEKNERVKK 714

RESULT 11
 BAG_STRAG STANDARD; PRT; 1164 AA.
 AC 021951;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE IGA FC receptor precursor (Beta antigen) (B antigen).
 GN BAG
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1311;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
 RC STRAIN=LA239;
 RX MEDLINE=91312121; PubMed=1857207;
 RA Jeristroem P.G., Chhatwal G.S., Timmis K.N.;
 RT "The Iga-binding beta antigen of the c protein complex of Group B
 RT streptococci: sequence determination of its gene and detection of two
 RT binding regions";
 RL Mol. Microbiol. 5:843-849(1991).
 RN (2)
 RP IDENTIFICATION OF IG-LIKE DOMAIN.
 RX MEDLINE=97035265; PubMed=8880921;
 RA Bateman A., Eddy S.R., Chothia C.;
 RT "Members of the immunoglobulin superfamily in bacteria";
 RL Protein Sci. 5:1939-1942(1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (potential).
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X59771; CAA42442.1; ..
 DR PIR; S15330; FCSOAG.
 DR InterPro; IPR004829; Csurface antigen.
 DR InterPro; IPR005877; Gpos YSIRK.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003599; Ig.

InterPro; IPR006192; LPXTG.
 Pfam; PF00746; Gram_pos_anchor; 1.
 Pfam; PF05062; RICH; 1.
 Pfam; PF04650; YSIRK signal; 1.
 ProDom; PD153432; Csurface_antigen; 1.
 SMART; SM00409; IG; 1.
 TIGRFAMS; TIGR01167; LPXTG anchor; 1.
 TIGRFAMS; TIGR01168; YSIRK signal; 1.
 PROSITE; PS0847; GRAM_POS_ANCHORING; 1.
 PROSITE; PS0835; IG LIKE; FALSE NEG.
 Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
 Immunoglobulin domain.
 FT SIGNAL 1 37
 FT CHAIN 38 1135 IGA FC RECEPTOR.
 FT PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 434 534 IG-LIKE.
 FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
 FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).
 FT DOMAIN 827 945 PRO-RICH REPEATS.
 FT SITE 1132 1136 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 FT SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;
 Query Match 7.2%; Score 172.5; DB 1; Length 1164;
 Best Local Similarity 21.6%; Pred. No. 0.3; 189; Indels 185; Gaps 25;
 Matches 124; Conservative 77; Mismatches 77

QY 36 LSAERPLNE-----QIAEAEDKIKTYPPENKPGQSNYSFVDNLNLKAIT-EKEK 86
 Db 149 LELENQFNETNRLHLIKQHEEVEKDKKAK---QOKTLKQSDTKVDLSNIDKELNHQSQ 204
 QY 87 IEK--ERQSISSPLDNKL-----NVEDVDSTKNRKLIDDYDSTKSGDLHK 130
 Db 205 VEKMAEQGITNEDKDSMLKKIEDIRKQAOQADKEDAEVKVREBELGLFSSTKAGLDQE 264
 QY 131 FQDDPDGLHOLDGTPLTAEDIVHKI-----AARIYEENDRAVDFDKIVSKLLNLG 179
 Db 265 IQE-----HVKKET--SSEENTQKDEHYANSLQNLAKSLEELDATTNQATQVKNQF 317
 QY 180 L-----ITESQAHLEDEVAEV-----LQKLISKE---ANNYEE 210
 Db 318 LENAQKKEIQPLIKETNVKLYKAMSESELOVEKELKHNSANLEDLVAKSEIVREYEG 377
 QY 211 DNKPTSWTE-----NOAGKIPEKVTMAAIODGLAKGENDETVSN 251
 Db 378 KLNQKNLPELQLEEEAHSKLVQVEDFRKFKTSQVTPKRVKGRDLAANENNQ---Q 434
 QY 252 TLTLNGLERTKTYSEDNPEELOVFPNFYALLKSIDSEKAEKETLITIMTKLIDFVK 311
 Db 435 KIELTVSPENITVYEGED-----VKFTVTAKS-DS-----KTLDFSD 471
 QY 312 MMVKYG-TISPBEVSVYLENLD-----EMIALQTK-----NKLEKNATD 349
 Db 472 LLTKYNPSVSDRISTNYKNTNDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKFTI 531
 QY 350 NISKLFPAPSKSHEETDSTKEEAA-----KMEKVGSLKD 385
 Db 532 TVQKKEKQVPTQKDSKTEEKVQPEKSNDKVQLQELIKSAQOELEKLEKAKELME 591
 QY 386 STKDDNSNPGSKTDPEKGTAYLEAIRKNIEMLKK-----HDKKGNKEDYDLSKMRDFIN 441
 Db 592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKIIGDSSSSSKYTYHYFNKYKSDFMN 648
 QY 442 KQADAYVEKGLDKEEAE-----AIKRIYSS 467
 Db 649 YQLHAQME--MLTRKVVQVMNKPVDNAEIKKIFES 681

RESULT 12
 TRDN CANFA
 ID TRDN CANFA STANDARD; PRT; 700 AA.
 AC P82179;
 DT 28-FEB-2003 (Rel. 41, Created)

QY	83	EKEKTEKRQSRISPLDNKLVNEDVDSTNRKLIDDDYSTSGSLGHFKQDDPDLGHLQDLD	143
Db	182	HKEKTEKKEKP-----ETKTMAKEERKAATEEKIK--KEVKGKGQEKVKPTAAKVKEVQ	233
QY	143	GTPLTAEADIVHKIAARI--YEENDRAVFDTKIVSKLLNLGLITESOHTL-	189
Db	234	KTPPKAKEKEKETAAVAHGEOQOYAFCTRTMDMFVHGDLPQGSPALPPPLPTVOASR	293
QY	190	-----EDEVAEVLQLKLSKEANNYEEDPNKPTS-----WTENQAQKIPE---	228
Db	294	PTPASPPTLEGKEBEKKAEKVVSETKKKEDEVKKSDKDTAIDVEKKEPGKAPETQK	353
QY	229	---KYTPMAAQTODGLAKGENDETNSVTLTTLNGLERTKTYTSDSNFEELOYPFNFALLK	285
Db	354	GTIKVVAQAAA- ---KDEKKEDSKTKT--PVEEHPKGGKQEKKE-----K	394
QY	286	SIDSSEKAKEKETLITIMKTLLIDFKVMVXYGTISPPEGVSYLENDLMIALOTKNKLEK	345
Db	395	VYEPKASKKHS-----APSE-----KQVKATERAKEE	424
QY	346	NATDNISKLFPAPEKSHEETDSTKEEAAMKEKEYGSLKSDTKDDNSNPGGKTDE--PKG	403
Db	425	TSAASTKAVPG---KKEEKTTVGOEIRKEK---SGKTSTASKDKPEIKDEKMPKA	478
QY	404	KTEAYLEAIRKNIBWLKHD---KKGKNEDYDISKMRDFIN-----KQADAYVEKG	451
Db	479	DKEVPKPPQSQVKKEESQVKKAEKPEQDIAPKPEKTVSHGKBPEKVVVKVKA--TEKA	537
QY	452	IILDK 455	
Db	538	AIEK 541	
 RESULT 13			
KF5C HUMAN			
ID	KF5C HUMAN	STANDARD; PRT; 957 AA.	
AC	O60282; O95079;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific		
DE	2).		
GN	KIF5C OR NKHC2 OR KIAA0531.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=98290545; PubMed=9628581;		
RA	Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,		
RA	Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. IX.		
RT	The complete sequences of 100 new cDNA clones from brain which can		
RT	code for large proteins in vitro."		
RL	DNA Res. 5:31-39(1998).		
RN	[2]		
RP	SEQUENCE OF 355-585 FROM N.A.		
RA	Engelender S., Sharp A.H., Colomer V., Tokito M.K., Lanahan A.,		
RA	Worley P., Holzbaur E.L.F., Ross C.A.;		
RT	"Huntingtin associated protein 1 (HAP1) interacts with the p150Glued		
RT	subunit of dynactin."		
RL	Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.		
-!	FUNCTION: KINESIN IS A MICROFILAMENT-ASSOCIATED FORCE-PRODUCING		
CC	PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.		
-!	SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT		
CC	CHAINS.		
-!	TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, PROSTATE, AND		
CC	TESTIS, AND MODERATE EXPRESSION IN KIDNEY, SMALL INTESTINE, AND		
CC	OVARY.		
-!	DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-		
CC	TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF		

CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
 CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
 CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
 CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
 CC VESICLES AND MEMBRANOUS ORGANELLES.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: AB011103; BAA35457.1; -;
 CC EMBL: AF010146; AAD01436.1; -;
 CC HSSP: P56536; 2KIN.
 CC Genew; HGNC:6325; KIF5C.
 CC MIM; 604593; -;
 CC GO; GO:0005871; C:kinesin complex; TAS.
 CC GO; GO:0003777; F:microtubule motor activity; TAS.
 CC GO; GO:0006996; P:organelle organization and biogenesis; TAS.
 CC InterPro; IPR001752; Kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC PRINTS; PRO0380; KINESINHEAVY.
 CC SMART; SM00129; KISG; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; microtubules; ATP-binding; Coiled coil.
 CC DOMAIN 1 386 KINESIN-MOTOR (BY SIMILARITY).
 CC DOMAIN 406 923 COILED COIL.
 CC DOMAIN 859 956 GLOBULAR.
 CC DOMAIN 174 315 MICROTUBULE-BINDING.
 CC NP_BIND 86 93 ATP (BY SIMILARITY).
 CC FT CONFLICT 355 360 TLKNI -> STHASV (IN REF. 2).
 CC FT CONFLICT 583 585 EFT -> DRV (IN REF. 2).
 CC SEQUENCE 957 AA; 109494 MW; A9F25BB1C994322A CRC64;

Query Match 7.0%; Score 167.5; DB 1; Length 957;
 Best Local Similarity 20.3%; Pred. No. 0.42;

Matches 103; Conservative 98; Mismatches 182; Indels 125; Gaps 23;

Qy 25 GSQDKSLNR-----ELSAERPLNEQIAEAEDEKIKKTYPPENKPGOSYFVDNLNLL- 78
 Db 321 GQRAKTIKTVSNVLELTAE-----EWKKYKEKEKNKTLKNVIQHLMEEL 367
 Qy 79 -----KAITEKEKIEKERSIRSSPLDNKLVEDVD-----STKNRKLIDYDSTKSG 126
 Db 368 NRWNGEAVPEDEQISAKDQK-NLEPCDNPIDNAPVVGISTBEK---EKYDEEIS 423
 Qy 127 LDHKFQDDPDLHLQDGTPLTAEIVHKIAARIYEEND-----RAVFDKIVSKLNLGLI 181
 Db 424 LYRQLDDKDEINQ-----QSQAELKQQLMDQDDELLASTRDYKEIKQELTFLQTE 476
 Qy 182 TSEQATLEDEVAELVKLSKEANNYE-----EDNKPTSTWENQAKIKPEKVPMAAIQD 238
 Db 477 NEA-----AKDEVKEVLQAL-EELAVNYDQKSQVEDKTRANEQLTDELAQKTTTLTTQ 531
 Qy 239 GLAK-----GENDETVSNTLTITNGLERR-----TK 264
 Db 532 ELSQQLSNHKKRATETIINLLKDLGEGGIIGND---VKTLADVNGVIEEFTMAR 588
 Qy 265 TYSEDNFEELQYFPNFALLKS---IDSEKAKEKETLITIMKTLDIFVQVMKYGTISPE 322
 Db 589 LYISKMKSEKVLNRSKQLESQMSDNKQNASERELAAQQLLISQHEAKISLT----- 644
 Qy 323 EGVSYLENIDEMALQTKNLEK---NATDNTSKLPAPSEKSHETDSTKE----- 371
 Db 645 ---DYQNMWME-----QKRQLESQDSLSLEELAKL---RAQERKMHEVSFQDKEHLRLIQ 694
 Qy 372 EAAKMEKEYGSLKDDSTKDDSNPGGKTDPKPKTEAYLEAIRKNTLEWLKHKDKGNKEDY 431

Db 695 DAEMKKALEQQWESHREAHQKLSRLDRDEIEBKIKIIDEIRDLNKLQLEQEK-LSSDY 753
 Qy 432 DLSKMRDFINKQADAYVEKGL--DKEE 457
 Db 754 NKLKIED---QEREMKLEKLLNDKRE 778

RESULT 14

MLP1_YEAST
 ID MLP1_YEAST STANDARD; PRT; 1875 AA.
 AC Q02455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 18-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MLP1.
 GN MLP1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=93247549; PubMed=8483450;
 RT Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RL "A new yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
 RT new open reading frames.";
 RL Yeast 9:1349-1354 (1993).
 CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
 CC REPAIR.
 CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----

CC EMBL; L01992; AAA34783.1; -;
 CC EMBL; X73541; CAA51948.1; -;
 CC EMBL; Z28320; CAA82174.1; -;
 CC PIR; S38173; S38173.
 CC SGD; S0001803; MLP1.
 CC DR GO; GO:0005635; C:nuclear membrane; IDA.
 CC DR GO; GO:0005654; C:nucleoplasm; IDA.
 CC DR GO; GO:0006006; P:protein-nucleus import; IDA.
 CC KW Coiled coil; DNA repair.
 CC FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 CC FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 CC FT CONFLICT 301 301 R -> A (IN REF. 1).
 CC SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 6.9%; Score 166.5; DB 1; Length 1875;
 Best Local Similarity 23.6%; Pred. No. 1;

Matches 112; Conservative 85; Mismatches 178; Indels 99; Gaps 28;

Qy 30 SLHNRLESAERPLNEQIAEAEDEKIKKTYPPENKPGOSYFVDNLNLLKAITE-KEKIE 88

Db 588 TLKSEKMDLESRIEELQKELEE--LKTSVPNE-----ASYSNV-----TIKQITETKRDLE 637

Qy 462 KRIYSL 468
Db 1596 KRVQAL 1602

Search completed: January 2, 2004, 12:34:53
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 40 Seconds
(without alignments)
3019.215 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGWTWLVLPVLPQAF.....EKGILDKKEARAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23.:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2062	86.0	466	11 Q8R1D7	Q8R1D7 mus musculus
2	1216	50.7	457	13 Q91847	Q91847 xenopus lae
3	198	8.3	7210	5 Q9V7G8	Q9V7G8 drosophila
4	198	8.3	9270	5 Q8MLD9	Q8MLD9 drosophila
5	197	8.2	2081	10 Q9LH98	Q9LH98 arabidopsis
6	196.5	8.2	1175	16 Q8XNW6	Q8XNW6 clostridium
7	196	8.2	2139	5 Q07569	Q07569 entamoeba h
8	191.5	8.0	2760	5 Q815Y2	Q815Y2 plasmodium
9	189	7.9	1434	5 Q81492	Q81492 plasmodium
10	188.5	7.9	1661	5 Q06166	Q06166 plasmodium
11	184	7.7	951	5 Q96229	Q96229 plasmodium
12	183.5	7.6	1790	3 Q07380	Q07380 saccharomyc
13	183.5	7.6	2612	5 Q815X5	Q815X5 plasmodium
14	180.5	7.5	3504	5 Q81L45	Q81L45 plasmodium
15	179.5	7.5	1320	11 Q9JK25	Q9JK25 rattus norv
16	179.5	7.5	1510	5 Q25920	Q25920 plasmodium

17	179	7.5	853	10 Q9LFE4	Q9LFE4 arabidopsis
18	179	7.5	1558	5 Q96275	Q96275 plasmodium
19	179	7.5	1786	5 Q9U0P0	Q9U0P0 plasmodium
20	179	7.5	2269	5 Q26223	Q26223 plasmodium
21	179	7.5	2747	5 Q9BJX9	Q9BJX9 plasmodium
22	178.5	7.4	1323	5 Q9NB35	Q9NB35 plasmodium
23	178.5	7.4	3317	16 Q8EWP8	Q8EWP8 mycoplasma
24	178	7.4	840	5 Q815X4	Q815X4 plasmodium
25	177.5	7.4	1464	5 Q81IF6	Q81IF6 plasmodium
26	177	7.4	1112	16 Q8EWQ1	Q8EWQ1 mycoplasma
27	177	7.4	1871	10 Q9SRD5	Q9SRD5 arabidopsis
28	176.5	7.4	682	5 Q81BV6	Q81BV6 plasmodium
29	175.5	7.3	1964	5 Q8SWQ7	Q8SWQ7 loligo peal
30	175	7.3	10578	5 Q81SF5	Q81SF5 caenorhabdi
31	175	7.3	18519	5 Q81SF6	Q81SF6 caenorhabdi
32	175	7.3	18534	5 Q81SF7	Q81SF7 caenorhabdi
33	174.5	7.3	924	5 Q15738	Q15738 dictyosteli
34	174.5	7.3	1263	5 Q81K49	Q81K49 plasmodium
35	173.5	7.2	1795	5 Q81E35	Q81E35 plasmodium
36	172.5	7.2	1134	2 Q99051	Q99051 streptococ
37	172.5	7.2	1979	5 Q96133	Q96133 plasmodium
38	172	7.2	495	5 Q81EK6	Q81EK6 plasmodium
39	171.5	7.1	2083	5 Q9N435	Q9N435 caenorhabdi
40	171.5	7.1	3484	5 P91257	P91257 caenorhabdi
41	171.5	7.1	5507	5 Q81HN3	Q81HN3 plasmodium
42	170.5	7.1	2771	5 Q26216	Q26216 plasmodium
43	170	7.1	1140	4 Q9UL85	Q9UL85 homo sapien
44	170	7.1	1804	11 Q921Q1	Q921Q1 mesocricetu
45	170	7.1	3130	5 Q9BK46	Q9BK46 plasmodium

ALIGNMENTS

RESULT 1

Q8R1D7	PRELIMINARY;	PRT;	466 AA.
AC	Q8R1D7;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Similar to secretogranin III.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Eye;		
RA	Strausberg R.;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC024785; AAH24785.1; -		
SQ	SEQUENCE 466 AA; 52783 MW; AB1CB9667276315C CRC64;		

Query Match 86.0%; Score 2062; DB 11; Length 466;
Best Local Similarity 87.0%; Pred. No. 2.1e-94;
Matches 410; Conservative 17; Mismatches 36; Indels 8; Gaps 2;

Qy	1	MGFLGTGWTWLVLPVLPQAF-----PIQAPKPGSQSKSLHNRELSAERPLNEQIAEAEDTKTKT	57
Db	1	MGFLWTGWSLVLPVLSNGPIQAPFKPGSQSKSLHNRELSAERPLNEQIAEAADKIKKA	60
Qy	58	YPPEKPGQSNYSVDNLNLLKAITEKEKTEKQRSIRSSPLDNKLVNVEDVDSVTKNRKLI	117
Db	61	FPSEKSESNYSVDNLNLLRAITEKETVEKEQRSIRSPFPDQLNVEDVDSVTKNRKLI	120
Qy	118	DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLN	177
Db	121	DEYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVDFKIVSKLN	180
Qy	178	LGLITESQAHTLEDEVAEVLQKLISKANNYEEDPNKPTSWTENQAQKIPKVTPTMAAIQ	237
Db	181	LGLITESQAHTLEDEVAEALQKLISKANNYEETLDKPTSTENQDQGIKPEKVPFVAVQ	240

QY 238 DGLAKGENDTVSNTLTITNGLERRTKTYSBDNFEELQYFNFYALLKSIDSEKEAKE 297
 DB 241 DQFTNRENDTVSNTLTITNSGLERTNPHRDDFEELQYFNFYALLTIDSEKEAKE 300
 QY 298 TLITIMKTLDIFVKMKVYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPFA 357
 DB 301 TLITIMKTLDIFVKMKVYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPFA 357
 QY 358 PSEKSHETDSTKEAAKMEKEYGSLKDTKODNSNPGKTDPEKGTETAYLEAIRKNIE 417
 DB 356 PPEKSEETDSTKEAAKMEKEYGSLKDTKODNSNPGKTDPEKGTETAYLEAIRKNIE 415
 QY 418 WLKHKHKKGNKEDYDLSKMRDPIFKQADYVYKGLDKEAEAIKRIYSSL 468
 DB 416 WLKHKHKKGNKEDYDLSKMRDPIFKQADYVYKGLDKEAEAIKRIYSSL 466

RESULT 2
 Q91847
 ID Q91847 PRELIMINARY; PRT; 457 AA.
 AC Q91847
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Secretogranin III.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Int. pituitary;
 RX MEDLINE=96217355; PubMed=8632145;
 RA Holthuis J.C.M.; Martens G.J.M.;
 RT "The neuroendocrine proteins secretogranin II and III are regionally
 conserved and coordinately expressed with proopiomelanocortin in
 RT Xenopus intermediate pituitary."
 RL J. Neurochem. 66:2248-2256(1996).
 DR EMBL: X92872; CAA63478.1;
 SQ SEQUENCE 457 AA; 51903 MW; B3096512D4B076D7 CRC64;

Query Match
 Best Local Similarity 50.7%; Score 1216; DB 13; Length 457;
 Matches 269; Conservative 63; Mismatches 106; Indels 34; Gaps 10;
 QY 8 TWILVLPIQAPKPGSGODKSLHNRSLAEPLNEQIAEEDKIKTYPPENKPGS 67
 DB 9 TAIGFVHVQGFNPSIKQDKDVHNRSEERPLEEQIAEA--DTVSRGTSTENQEMLK 66
 QY 68 NYSFVONLNLKATKEKIEKERQSISSPFLONKLNVEDVSTKNRKLIDDYDSTKSL 127
 DB 67 NDSFADDLGLMSNGEKQGVKSGSTRGSGNRTGLDDASTKNLKLAEVFTTKSWP 126
 QY 128 DHKQDDPDGLHOLDGTPLEADIVHKIARIYEENDRAVDFKIVSKLNLGLITESQAH 187
 DB 127 DYKYEDPDGLHOLD-APFTAEDIVRKIATRIYEENDRGVDFKIVSKLNLGLITESQAY 185
 QY 188 TLDEVAEVLQKLSKEANNYEE-----DPNKTSWTENQAGKIPKVTFMMAIQG 239
 DB 186 TLDEVAEVLQQLIAEAKNREAGDLDSVRSVDYDDKREGM-----ET 234
 QY 240 LAKGENDT--VSNTLTITNGLERRTKTYSBDNFEELQYFNFYALLKSIDSEKEAKE 297
 DB 235 LNKNEDESSETKNEDSLST--VERRNELSPED---DLQYFNFYALLKSIDSEKEAKE 290
 QY 298 TLITIMKTLDIFVKMKVYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPFA 357
 DB 291 TLITIMKTLDIFVKMKVYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPFA 347
 QY 358 PSEKSHETDSTKEAAKMEKEYGSLKDTKODNSNPGKTDPEKGTETAYLEAIRKNIE 416
 DB 356 PPEKSEETDSTKEAAKMEKEYGSLKDTKODNSNPGKTDPEKGTETAYLEAIRKNIE 415

DB 348 PSEKSHETDSTKEAAKMEKEYGSLKDTKODNSNPGKTDPEKGTETAYLEAIRKNIE 405
 QY 417 EWLKHKHKKGNKEDYDLSKMRDPIFKQADYVYKGLDKEAEAIKRIYSSL 468
 DB 406 EWLKHKHKKGNKEDYDLSKMRDPIFKQADYVYKGLDKEAEAIKRIYSSL 457

RESULT 3
 Q9V7G8
 ID Q9V7G8 PRELIMINARY; PRT; 7210 AA.
 AC Q9V7G8
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CG18255 protein.
 GN STRN-MLCK OR CG8304 OR CG18255.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burks J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Durbin K.J., Svangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Ye J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Buesam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorset V., Dou L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez C., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,

```
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards R., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mirra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tuzy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richer J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;
RT "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the ENBL/GenBank/DBJ databases.
DR EMBL; AE003808; AAF58087.2; -
DR FlyBase; FBgn0013988; Strin-Mlck.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG c2.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG; 13.
DR SMART; SM00409; IG; 21.
DR SMART; SM00408; IGC2; 20.
DR PROSITE; PS00835; IG-Like; 20.
DR PROSITE; PS00290; IG MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 6B2E7395C07140D2 CRC64;

Query Match      8.3%; Score 198; DB 5; Length 7210;
Best Local Similarity 23.8%; Pred. No. 0.55;
Matches 110; Conservative 83; Mismatches 180; Indels 90; Gaps 19;

Qy 35 ELSEPLNEIOAAREEDKIKKTYPPENKPGSQSVFVNLNLKALTEKEKIERQSI 94
Db 4757 EIVSEKITDEKAQSQKEVGS---EAKPKAKV-----LEKSEIEEKLEKDKKEQ 4806

Qy 95 RSSPLDKLVNVEDVSTKNRKLIDD--YDSTKSGLDHFKFDDPLHQLDGTPLTAEDIV 152
Db 4807 TESAIDEKSKQAEVSEIVSEKITDEKAQSQKEVKDS-EAKPKAKVLEKKSIEEEKLE 4865

Qy 153 HK----TAARIYENDRAVFKIVSKLNLGLITESTQHTLEDEVAEVQ-KLISKANN 207
Db 4866 NKKEKQTESAIDEKSKQAEVSEIVSEKITDEKAQSQKEVKDS-EAKPKAKVLEKKSIE 4925

Qy 208 YEEDPNKPTSTENQAGKIPKVTYPMALQDGLAGNDETNTVNTLTNGLERRTKYS 267
Db 4926 EEKLEKKEKQTE-----SAIDEKFOKAEVSEIVSEKIT-----DEKAESE 4967

Qy 268 EDNFEELQYFPNFYAL--KSIDSEK--EAEKET-----LITIMKTLID-- 308
Db 4968 KEEVKDEAKPKAKVLEKKSIEEEKLEKDEKTEKTESAIDEKSKQAEVSEIVSEKITDEK 5027

Qy 309 ---FVKMMVKYGTISP-----EGVSVYLENDEMIALQTKNKLKATNATNISKLFPPS 359
Db 5028 AQESQKEVKDSQAEKPKAKVLEKKSIEEEKLEKDEKQETKV--ATDTKSTQVSE 5084

Qy 360 EKSHEETDSTKEAAKMEKVGSKDSTKDNNSPGGKTDPKKTAYLEAIRKNLEWL 419
Db 5085 IVLEKISIEEAKESQKVE-----LKDS-----EAKSKAKVLEKKSITLKEKL 5126

Qy 420 KKHDKGKNKEDYDLSKMRDFINKQADAVEKGILDKKEAEAIK 462
```

```
Db 5127 DENDKK-QKEDGATNKSQK--AQAADVVPK--ISEEKVAEIK 5164

RESULT 4
Q8MLD9 PRELIMINARY; PRT; 9270 AA.
ID Q8MLD9
AC Q8MLD9;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
CG18255-PA.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MDLLine=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pittman G.S., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzales M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
```



```
Db 1009 --NREKKEYEKSKTBEAKKESQDKKREKDSERKSKKEBSRDLKAKKKEE 1066
Qy 401 PKGTEAYLEAIRKNIEWLKHDKGKNGEDYLSKQDFINQKQADYVVEKGILDKKE 457
Db 1067 TKEKES-----ENHKSCKKEDKHEHDNKSMMKEED--KKEKKKHEESKRKEE 1115

RESULT 6
ID Q8XNW6 PRELIMINARY; PRT; 1175 AA.
AC Q8XNW6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Probable exonuclease.
GN SBCC OR CPE0216.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
EX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
DR EMBL; AP003185; BAB79922.1; -.
DR InterPro; IPR003439; ABC_transporter.
KW Exonuclease; Complete proteome.
SQ SEQUENCE 1175 AA; 136878 MW; 7C6D2366525019C1 CRC64;

Query Match 8.2%; Score 196.5; DB 16; Length 1175;
Best Local Similarity 22.6%; Pred. No. 0.078;
Matches 116; Conservative 83; Mismatches 196; Indels 119; Gaps 21;

Qy 29 KSLHRELASRLNEQIAEAEDEKIKKTPPENKPGOSYFVDNL-NLLKAITEKE-- 85
Db 256 KELYDKRIEE-----SLVSRSEIKSFERVEISNKADKIVFINNLEILKEINKEDLK 311
Qy 86 --KIEKQSTRSPDLNKLNVEDVSTNKNKLIDDYSTKSGLDHKFPDDPGHQLDGL 143
Db 312 FSELNKLLEELNREENKLFEBFTKKKEKLPDLRLKKEKLLS--QKRDILFQIKA 369
Qy 144 TPLTAEDIVHKI-----AARIYENDRAVPDKIVSK-----LLNLG 179
Db 370 DGVKLKEACKIFEDRSKCDTKLNSIENENKRLNEELKEEKEERKEELFVHEEFKNKINS 429
Qy 180 LITESQAHTLEDEVAEVLQKLI--SKEANNYEEDPNKPTSWTNOAGKIPKVTMAAIQ 237
Db 430 LFILNSYSLQKQNEIKSEVEELKYYIKNTEDKEKSEK-----DLKVVESLSKIR 482
Qy 238 DGLAK-----GENDTVSNTLTITNGLERKTKYSEDNFE--LQYF 277
Db 483 DKLESLLAKETPGDSNLSILEKQIKLGEYREKLNKYKEIKNSLESLKT--KNPFSEKLT 540
Qy 278 PNFYALLKSIDSEKAEKETLITIMKT--LIDFVMMVKYGTISPREGVSYL----- 329
Db 541 ENQKLL-----EKEVELKDYINKVVELAHLKRLNLEGECCPGVSGSTHHELNKVEK 595
Qy 330 -NLDE---MIALQTKNKLKNATDNISKL--FPAPSEK-----SHEETDSTKEBAK 375
Db 596 INLESNEKTIILLESKEKELILEFSKIEATLEYENKKEELNLSIEEVEGVNEERLK 655
Qy 376 -MEKEYGSLDKDSDNSNPGKTDPEPKGTAYLEAIRKNIEWLKHDKGKNGEDYLS 434
Db 656 FLEEFNTLKDIEEFNLK-----ENLEKDLKXK--EKKNNLE----- 693
Qy 435 KQRFINKQADYVVEKGILDKKEEAEAKIYSS 468
Db 694 ---NIFNKAIEVILCEKIVREKEIAISKELDKEL 724
```

RESULT 7

```
Qy 007569 PRELIMINARY; PRT; 2139 AA.
AC Q07569; Q02504;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Myosin heavy chain.
GN MHCA.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
EX MEDLINE=93295430; PubMed=8515774;
RA Raymond-Denise A., Sansonetti P., Guillen N.;
RT "Identification and characterization of a myosin heavy chain gene
RT (mhca) from the human parasitic pathogen Entamoeba histolytica.";
RL Mol. Biochem. Parasitol. 59:123-131 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RA Guillen N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L03534; AAB48065.1; -.
DR HSP; P08799; INMD
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SW
SQ SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;

Query Match 8.2%; Score 196; DB 5; Length 2139;
Best Local Similarity 21.0%; Pred. No. 0.17;
Matches 107; Conservative 108; Mismatches 191; Indels 104; Gaps 23;

Qy 20 FPKPGSQDKSLHRELASRLNEQIAEAEDEKIKKTPPEN-----KPGOSYFV 72
Db 859 FEEGKKKDEIED-----LKKLAE---EIKKREAAENALASATAKTGLEAKIQ 906
Qy 73 D-----NLLKAITEKEKIEKERSIRSPSLNKLNVEDVD- 109
Db 907 DLEDKTISESKLSAAELDKQELNKLNTENLEEDKEELKTDNLKGLDKSLKGEDLEV 966
Qy 110 -----STKNRKLIDDYSTKSGLDHKFPDDPGHQLDGLPTAEDIVHKAARIY 160
Db 967 ETELNSQIINTLN-ATVDKDKTIAEMQESIDEKEDEITKLGDIKLEEEKDDL----- 1020
Qy 161 EENDR-----AVFDKIVSKLNLGLITISQAHTLEDEVAEVLQKLI-SKEANNYEEDPNKPT 216
Db 1021 -EQDRAVDSATKDDIAKLNKTI-----ECEDAKDEIAKLEQLEDEE--NKNKDLTNEL 1073
Qy 217 SWTENQAKPIPEKVTMAAI-----ODGLAG--ENDETVSNTLTITNG-LERRTKYS 267
Db 1074 QOTQLKLGETEKSALAAQVAATKASDERDTLSQNLNEKLTITKNLTKTADLEKKISGLK 1133
Qy 268 EDNFEELQYFPNFVALLKSID-----SEKAEKETLITIMKTILIDFV-KMMVKYGTISP 321
Db 1134 QD-YEDLEDDKN-----KIEGLRNQRKIKELDDIEITGADVSVLQKQKEEY----- 1181
Qy 322 EGVSYLENLDEMIALQTKNKLKNATDNISKLPAPSEKSHSEBTSTKEBAK-----ME 377
Db 1182 ESQIAKMQEKEKALGNDVKNK-EKTIKE--KELEIQSLQEKLDETEVEKEDAEKKKEIE 1238
Qy 378 KEYGSLDKDSDNSNPGKTDPEPKGTAYLEAIRKNIEWLKHDKGKNGEDYLD-SKM 436
Db 1239 KEMKALQEE--KENVESSKNSTEKDKKLEEDNLKDTQKKLDMDTADNEKLLAKAKDLQA 1297
```

```
Qy 437 RDKFQADAYVEKGLDKEAEAKRIYS 466
Db 1298 NEVDNHERAVADAELNKKQASDKELNS 1327

RESULT 8
Q815Y2
ID Q815Y2 PRELIMINARY; PRT; 2760 AA.
AC Q815Y2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL0315C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_TaxID=36329;
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014845; AAN36152.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2760 AA; 330896 MW; EE9964C4845181AC CRC64;

Query Match 8.0%; Score 191.5; DB 5; Length 2760;
Best Local Similarity 21.1%; Pred. No. 0.37;
Matches 121; Conservative 92; Mismatches 207; Indels 153; Gaps 24;

Qy 29 KSLNRLSARPLNEQAEAE-----DKIKTYPENKPGQSNYSFVD 73
Db 1452 KNEAKEEEETKNDQWNEKEMVNSKNERIVQVHNEIKNNKKEBEGKKNLKER 1511
Qy 74 NLN--LLKAITEKEKIERQ--SIRSSPLDNLNVEDVDSTKNRKLIDDYDSTKGLDH 129
Db 1512 EINDCLNDYINKQKKEKKNNWAMYGRPIVKRQNNRNINIKNDLKKLYSSKSESGFNDY 1571
Qy 130 KF-----QDDPDGLHLDGTPLTAE--DIVH-----KIAARIYEEN-- 163
Db 1572 AFVAERFEVITGYNSEFDYLSIDNOAKNEKNNDIIHNNIIKISKMKENIYENSFP 1631
Qy 164 ---DRAVDF-----KIVSKLNLGLITESQA----- 186
Db 1632 HTYGRPIYEKKSNPNYNIKISTHNAILKKRKKTLNKSISINSFTKNSNNKIVKR 1691
Qy 187 -----HTLEDEVAEVLQKLSKANNYEEDP-----NKPTSWTENQAGK 225
Db 1692 TSIKNNNTIDNNSTTKIIHKEQN--VEDQGYIDLTKRKLIDALDEINQOTQOKNLK 1749
Qy 226 IPEKVTPTMAAIQDLAKGENDETSTNLTITNG--LERATKYSEDNFELOYFFNFVAL 283
Db 1750 ITENIT-----QVQKHGQN-----VSNIIKTGAMLIETIKR-GNDNNEEDQFSBELKAL 1800
Qy 284 LKSIDSEKAKEKETITIMKTILDFVGMVVKYGTISPEEGVSYLENIDEM-----I 335
Db 1801 EK-LKKLKEKLTLEELTEE--BEKKLEELNSLKVEEKKKKNNKKEEMKMEEMKI 1856
Qy 336 ALQTKNLEK-----NATDNISKL-----FPAPSEKSHETDSTKEAAKM--EKEY 380
Db 1857 ELORKKEEKLQYRRQRIKRLBELKKELKLEENMKLEEQKREIKKMEEKER 1916

Qy 381 GSKLDSTKDDNSNPGKTDPPKGTAYL-----FAIRKNIE--WLKKHDKKGNKEDY 431
Db 1917 EKLKTIKEQKQKEEMKKQEQRKKEQIKLMEEMRKFEQKKVKLWLESEKREBELKKL 1976

Qy 432 DLSKMRDFINKQADAYVEKGLDKEAEAKRI 464
Db 1977 EQOKRE-ERKKMEBEKKREBEKKREBEIKKM 2008

RESULT 9
Q81492
ID Q81492 PRELIMINARY; PRT; 1434 AA.
AC Q81492;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Mature parasite-infected erythrocyte surface antigen (MESA) or
DE PfEMP2.
GN MESA.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_TaxID=36329;
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrall B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Horneby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51374.1; -.
SQ SEQUENCE 1434 AA; 168287 MW; AB0005F9DC26C989 CRC64;

Query Match 7.9%; Score 189; DB 5; Length 1434;
Best Local Similarity 21.1%; Pred. No. 0.23;
Matches 101; Conservative 96; Mismatches 180; Indels 102; Gaps 23;

Qy 27 QDKSLNRLSARPLNEQ-----IAEAEEDKIKTYPENKPGQSNYSFVDNLJLKA 81
Db 501 KDKVLGSGEDKEDKKNDEQKVLGEGKEDVK-----EKNDGKD-KVIGSEKTKEI 554
Qy 82 TEKEKIEK-----ROSIRSSPLDNLNVEDVDSTKNRKLIDDYDSTKGLDHKFDQ--- 133
Db 555 --KEKVEKRVKKCKKKVKKGIKEN--DTEGNDKVGKPEII--IEBVKEIKKQVEDGIK 608
Qy 134 --DPDGLHLDGTPLTAEEDIVHKAARIYE--ENDRAVPDKIVSKLNLGLITESQAH 188
Db 609 ENTEGNDKVGKPEITEEVKEIKKQVEEGIKENDETEGNDKVGKGP-----EIITEVKEE 664
Qy 189 LEDEVAEVLQ-----KLISKEANNYEEDPNKPTSWTENQAGKPEKVTPTMAAIQDGL 240
Db 665 IKKQVEEGIKENDTESKDKLIGQEIITEEVKEIKENDETNK-----DKVIGQEIITEV 719
Qy 241 AKG--ENDETVSNLTITNGLERRTYSEDNFELOYFFNFYALLKSIDSEKAKEKET 298
```

```

Db 720 KEGIKEND-----TENKDVIGQBIITE-----EYKKEIEQKEKNGKEN 759
QY 299 LITI-----MKTLDIFVQMVVYGTISPBEQV--SYLENLD-----EMIALQTKNKLK 345
Db 760 ILEIDIVIGQVIEEVKVKK---KVEKIGKENHTSEKDKVIGQBIIVVEEVEKEIEK 816
QY 346 NATDNISKLPAPSEKSHETDSTKEAAKMEKEVGSLSKDTKODNSNPGKTDPEPKGT 405
Db 817 QVEGIGK-----ENDTESKDVIGQVIGK-----DVNEBGFENKDKVTKQEKVK- 861
QY 406 EAYLEAIRNRNIEWLKHKHKGKEDYDLSK--MRDFINKQADAYVEKGILDKERAEALK 462
Db 862 EYKVEKVKVKVKKRNNKRNKRNKONVIGKEIMKEDVNEKDTANKDKIEQKEKEEYK 920

RESULT 10
Q06166
ID Q06166 PRELIMINARY; PRT; 1661 AA.
AC Q06166;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Mature PARASITE-infected erythrocyte surface antigen (Antigenic
protein PREMP2).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RX MEDLINE=92158014; PubMed=1741020;
RA Coppel R.L.;
RT "Repeat structures in a Plasmodium falciparum protein (MESA) that
binds human erythrocyte protein 4.1."
RL Mol. Biochem. Parasitol. 50:335-347(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RX MEDLINE=93122844; PubMed=1478701;
RA Saul A., Vegganesh F., Howard R.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 797-850 FROM N.A.
RX MEDLINE=93122844; PubMed=1478701;
RA Saul A., Vegganesh F., Howard R.J.;
RT "Conservation of repeating structures in the PfEMP2/MESA protein of
Plasmodium falciparum."
RL Immunol. Cell Biol. 70:353-355(1992).
DR EMBL; AF056936; AAC13303.1; -.
DR EMBL; S52458; AAB24869.1; -.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00076; DnaJ_2; 1.
KW Antigen.
SQ SEQUENCE 1661 AA; 195479 MW; AF340527D85A9D29 CRC64;

Query Match 7.9%; Score 188.5; DB 5; Length 1661;
Best Local Similarity 19.9%; Pred. No. 0.29;
Matches 94; Conservative 90; Mismatches 189; Indels 99; Gaps 15;

QY 22 KPGSQDKSLNRELNSAERPLNEQIAEAEEDKTKYPPENKPGQSNYSFVDNLLKAI 81
Db 895 KDTANKDKIEQKEKEEVEKEEVEKEEVEKEE-----V 937

QY 82 TEKEIEKERSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSLDHLKFDQDPGLHQL 141
Db 938 KEKEVEKSEKEEVEKEEVEKEEVEKEEVEKEEVEKEEVEKEEVEKEEVEKEEVEKEE 996

QY 142 DGTPLTAEDI VHKAARYEEN-----DRAVFDKIVSKLNLGLITSEQAHTLEDEVAE 195
Db 997 IGOEIIIEIKVEKVKRNNKRNKRNKONVIGQIEMNEKDTANKDKVIEQKEKEE 1056

QY 196 VLQKLISKANNYEDPNKPTSWTENQAGKIPEKVTPTMAAIQDGLAGENDETVNTLTL 255

```

```

Db 1057 VKEEVEKEEVEKE-----KBEVK-----EKEEV----- 1081
QY 256 TNGLERTKTVSEDNPEELQVFPNFYALLKSIDSEKAEKKEK-----LITIMKTLIDFVK 311
Db 1082 -----KEKDTESKDKIEQ-----KEKEVEKVEKEDTENKDKVIGQBIIEIK 1127
QY 312 MMVKYGTISPBEQVSYLENLDSEMIALQTKV--LEKNATDNISKLPAPSEKSH--EETDST 369
Db 1128 KEVKRYV---KRRNNKRNKRNKONVIGQIEMNEKDTANKDKVIEQKEKEEVEKEEVE 1184
QY 370 KEEAAKMEKVGSLKDS-----TKDDNSNPGKTDPEPKGT---EAYLEAIRKN 415
Db 1185 KEKEVEKEEVEKEDTESKDKIEQKEEVEKEEVEKEDTENKDKVIGQBIIEIKKE 1244
QY 416 IE--WLKHKHKGKEDYDLSK--MRDFINKQADAYVEKGILDKERAEALKRI 464
Db 1245 VKKRVKRNKRNKRNKONVIGQIEMNEKDTANKDKVIEQKEKEEVEKEEVE 1296

RESULT 11
O96229
ID O96229 PRELIMINARY; PRT; 951 AA.
AC O96229;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF0680W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum."
RL Science 282:1126-1132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
Fraser C.M., Barrell B.;
RA "Genome sequence of the human malaria parasite Plasmodium
falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE001410; AAC71925.2; -.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; ACBD889358A84F4P CRC64;

Query Match 7.7%; Score 184; DB 5; Length 951;
Best Local Similarity 22.0%; Pred. No. 0.25;
Matches 104; Conservative 76; Mismatches 182; Indels 110; Gaps 17;

QY 25 GSQDKSLNRELNSAERPLNEQIAEAEEDKTKYPPENKPGQSNYSFVDNLLKAI 84
Db 161 GKQDINSNAE--NKKDVKEGVEKEEKEEKEEKEEKEEKEE-----DD 206
QY 85 EKIEKERSIRSSPLDNKLN----VEDVDSTKNRKLIDDYDSTKSLDHLKFDQDPGLHQL 140

```

```

Db 207 HKVEENKSDHVKVEENKSDHDKIEVKKVVEHEDEDEKKEKSEKKNKDNKEND 266
QY 141 LDGTPPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGLITSESQHTLEDVAEVLQKL 200
Db 267 EDNDEISDEVEDDDVEEDKQNDIDDDK-----KETDKHLEBENBIEKE 315
QY 201 IS-KEANNYBEDPNKTSW--TENQAGKIPKVTPTMAAIQDGLAGNDETVSNTLITLNG 258
Db 316 FSDKKXGKGNKDKTEKSKDTEKSKDIEK-----EKSQKKEKES----- 357
QY 259 LEERTKYSEDNPEELQYFNFVALLKSIDSEKAEKETLITIMKTILIDFVKMMVYGT 318
Db 358 -KDEKEKKGKKEKE-----KSKDIEKE-KKQKQIEKSK-DTAKEKEDKD 403
QY 319 ISPEEGVSYLENDEMIALQTKNLEKKNATDNISKLPAPSEKSH-----E 364
Db 404 IEKEKS-----KQMEKLNKQNEKKKQDDNEKK-----KNDKQDHDNDNDMEBIE 452
QY 365 ETDSTKEBAKME-----KEYGSLKDKSTK-DDNSNPGGKTDBPKGTEAYLEAI 412
Db 453 ENDEDEDEDEKKNKKNKNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 512
QY 413 RNIEWLKKHDKKNGKEDYLSKMDP-----INKQADAVVEKGILD 454
Db 513 NENNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 564

```

RESULT 12

```

Q07380 ID Q07380 PRELIMINARY; PRT; 1790 AA.
AC DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical 206.5 kDa protein YDL058W.
GN US01 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecher H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74106; CAA98621.1; -.
DR EMBL; Z74105; CAA98620.1; -.
DR SGD; S0002216; US01.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol_p115_C.
DR Pfam; PF04871; Usol_p115_C; 1_head.
DR Pfam; PF04869; Usol_p115_head; 1.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;

```

```

Query Match
Best Local Similarity 20.3%; Score 183.5; DB 3; Length 1790;
Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;
QY 29 KSLHN--RELSARPLNEQIAEAEEDKIKKTPPENKPGQSNYSFVDNLNLKAI-----T 82
Db 965 KSLANNYKQMAE---NESLIKAVEE-----SKNESSTQLSNLQNKIDMSQ 1008
QY 83 EKEKIEKQRSTSSPLDNKLNVEDSTKNRKL I-----DDYSTKGLDHKQD--- 133
Db 1009 EKENFQIRGSIEKNEQLKKTISLEQTKHEIISKSDSSKDEYSQISLKEKLETATT 1068
QY 134 -DPDGLHQDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGLITESQAH----- 187
Db 1069 ANDENVNKISLTKTKEELEALAA--YKNLKNLETKLETSEKALKEVKNESHKKEK 1126
QY 188 -TLEDEVAEVLQKLISKAN-----NVEED--PNKPTSWTENQAKIPEK 229

```

```

Db 1127 IOLEKEATETKQOLNSLRANLESLEKEHEDLAAQLKKYBEQIANKEROYNE--EISQLNDE 1185
QY 230 VTPMAAIQDGLAGNDETVSNTLITLNGLEERT-----KTYSEDN--- 270
Db 1186 ITSTQENESIKK-KNDELEGEVKAAMKSTSEQSNLKKSEIDALNLQIKELKKKQNTNEA 1244
QY 271 -----PEELQYFNF-----YALLKSIDSEKAEK 294
Db 1245 SLLESIKSVESVETVKIKELQDECNFKPEKVESELEDKLGASEDKNSKYLELQK--ESEKIKE 1303
QY 295 EKETLITIMKTILIDFVKMMVK----- 315
Db 1304 ELDAKTEIKIQLKIEKITNLSKAKESSELSRLKKTSEERKNVABEQLEKLNKNEIQIKNQ 1363
QY 316 -----YGTISPE--EGVSYLENDEMIALQTKNLEKKNATDNISKLPAPSEK 361
Db 1364 AFEKERKLLNEGSSITQYSEKINTLE--DLIRLQNEENELKAKEIDNTRSELEKVSLS 1421
QY 362 SHEETDSTKEBAKMEKEVGSUKDS--TKDDNSNPGGKTDEP-----KGKTEAYLEAIRK 414
Db 1422 NDELLEEKQNTIKSLQDEILSYDKITRNDEKLLSIERDNKRDLSLKEQLRAAQESKAK 1481
QY 415 NIEWLKKHDKKNGKEDYLSKMRDPINKQADAVVEKGILD--KEEAATIKR 463
Db 1482 VEEGLKLEESSEKAELEKSEKMM--KLESTIESNETELKSSMETIRK 1530

```

RESULT 13

```

Q815X5 ID Q815X5 PRELIMINARY; PRT; 2612 AA.
AC DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFLO350C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Petrea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014845; AANJ36159.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2612 AA; 311421 MW; C950B6E466E3D1C5 CRC64;

```

```

Query Match
Best Local Similarity 22.3%; Score 183.5; DB 5; Length 2612;
Matches 114; Conservative 98; Mismatches 196; Indels 103; Gaps 23;
QY 27 QKSLHNRNLSARPP---LNEQIAEAEEDKIKKTPPENKPGQSNYSFVDNLNLKAI 83
Db 991 REKEIDNRRNRIERKENFLNHTYKKELEEKI-----NNKKEKLEMLK--KE 1036
QY 84 KEKIEKERSIRSSPLDNKLNVE-----DVDSTKNR--KLIDDYSTKSG 126
Db 1037 LESLEKEKEKQIIECEYNNLQNEKEEIQNRNNLIKQKELNSRMDRYNELLEDLNKRKE 1096
QY 127 LD---HKFODDPG-----LHOLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNL 178

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 46 Seconds

(without alignments)
1614.870 Million cell updates/sec

Title: ABU66614

Perfect score: 2399

Sequence: 1 MGFLGTGTLVLVLPIQAF.....EKGILDKEEAEAKRIYSSL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2399	100.0	468	21	AA18926
2	2399	100.0	468	22	AAU12216
3	2399	100.0	468	22	ABG7600
4	2399	100.0	468	23	ABG95925
5	2399	100.0	468	24	ABU69097
6	2399	100.0	468	24	ABU69120
7	2399	100.0	468	24	ABU71580
8	2399	100.0	468	24	ABU72026
9	2399	100.0	468	24	ABU72183

10	2399	100.0	468	24	ABU66614	Human PRO polypept
11	2399	100.0	468	24	ABU66890	Human secreted/tra
12	2399	100.0	468	24	ABU59695	Novel secreted and
13	2399	100.0	468	24	ABG73287	Human neurosecret
14	2399	100.0	567	22	AAO13905	Human polypeptide
15	2392	99.7	468	22	AAW93910	Human polypeptide,
16	2384	99.4	468	21	AAW81955	Human HgIII prote
17	2271	94.7	470	22	ABG12976	Novel human diagno
18	2102.5	87.6	471	23	ABW57046	Mouse ischaemic co
19	1974	82.3	396	24	AAO16322	Human secreted pro
20	594	24.8	316	22	ABG14006	Novel human diagno
21	475	19.8	212	22	ABG14003	Novel human diagno
22	184	7.7	665	21	AA18278	Plasmodium falcipa
23	180.5	7.5	1881	23	ABF73809	Candida albicans e
24	179	7.5	1558	21	AA181324	Plasmodium falcipa
25	179	7.5	1786	18	AAW24790	P. falciparum live
26	179	7.5	1787	23	AAU96699	Plasmodium falcipa
27	178.5	7.4	1392	20	AYO66999	Restin protein seq
28	175.5	7.3	1427	12	AA10534	Human 160kD mediat
29	174.5	7.3	1164	19	AAW40537	Group B streptococ
30	173	7.2	1099	19	AAW40538	Mutant C-beta prot
31	172.5	7.2	1164	17	AAW85781	Group B Streptococ
32	172.5	7.2	1164	21	AAW84459	Amino acid sequenc
33	172.5	7.2	1979	21	AA18171	Plasmodium falcipa
34	171	7.1	1093	19	AAW40540	Mutant C-beta prot
35	170	7.1	1213	22	AAW40016	Human polypeptide
36	169.5	7.1	1788	22	ABW62995	Drosophila melanog
37	169	7.0	1129	21	AAW84462	Amino acid sequenc
38	168	7.0	1164	19	AAW40541	Mutant C-beta prot
39	168	7.0	1164	21	AAW84463	Amino acid sequenc
40	167	7.0	493	13	AAW26944	P.falciparum USA g
41	166.5	6.9	1135	21	AAW84460	Amino acid sequenc
42	166	6.9	1128	19	AAW40539	Mutant C-beta prot
43	166	6.9	1128	21	AAW84461	Amino acid sequenc
44	166	6.9	1192	21	AA18165	Plasmodium falcipa
45	166	6.9	2633	22	ABG06505	Novel human diagno

ALIGNMENTS

RESULT 1
AA18926
ID AA18926 standard; Protein; 468 AA.

XX AA18926;

XX 08-FEB-2001 (first entry)

XX A novel polypeptide designated PRO5990.

DE Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW PRO4424; PRO4430; PRO4430; PRO4499; tumour; obesity; diabetes;
KW insulinemia; kidney disorder; Bergers disease; nephropathy;
KW Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW Crohns disease.

XX Homo sapiens.

Key	Location/Qualifiers
FT Modified-site	68..71
FT Modified-site	/note= "N-glycosylation site"
FT Modified-site	70..73
FT Modified-site	/note= "casein kinase II phosphorylation site"
FT Modified-site	82..85
FT Modified-site	/note= "casein kinase II phosphorylation site"
FT Modified-site	97..100
FT Modified-site	/note= "casein kinase II phosphorylation site"
FT Modified-site	125..128
FT Modified-site	/note= "casein kinase II phosphorylation site"
FT Modified-site	143..148

PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX WPI; 2001-183260/18.
DR N-PSDB; AAF92132.
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping. -
XX Claim 12; Fig 150; 278pp; English.
XX The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.
XX SQ Sequence 468 AA;
Query Match 100.0%; Score 2399; DB 22; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGFLGTGFWILVLPQAFKPGGQSKLHNLRLSARPLNEQIAEAEEDKIKKTYPP 60
Db 1 MGFLGTGFWILVLPQAFKPGGQSKLHNLRLSARPLNEQIAEAEEDKIKKTYPP 60
Qy 61 ENKPGQSNYSFVDNLNLKATTEKIEKESIRSSPLDNKLNVEDVSTKNRKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLKATTEKIEKESIRSSPLDNKLNVEDVSTKNRKLIDDY 120
Qy 121 DSTKSGLDHKKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
Db 121 DSTKSGLDHKKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
Qy 181 ITESQHTLEDEVAEVLQKLSKEANNVEEDPNKTSWTENQAGKIPEKVTPTMAAIQDGL 240
Db 181 ITESQHTLEDEVAEVLQKLSKEANNVEEDPNKTSWTENQAGKIPEKVTPTMAAIQDGL 240
Qy 241 AKGENDETSTNLTITNGLERRTKYSNDNPELOYPFNFVALLKSIDSEKAEKETLI 300
Db 241 AKGENDETSTNLTITNGLERRTKYSNDNPELOYPFNFVALLKSIDSEKAEKETLI 300
Qy 301 TIMKTLLIDFVKMWVYGTISPPEGVSYLENLDEMIALQTKNKLKNATDNISKLPAPSE 360
Db 301 TIMKTLLIDFVKMWVYGTISPPEGVSYLENLDEMIALQTKNKLKNATDNISKLPAPSE 360
Qy 361 KSHETSTDEKAAKWEYGLSKDSTKDDNSNPGKGTDEPKGKTEAYLEAIRKNIEWLK 420
Db 361 KSHETSTDEKAAKWEYGLSKDSTKDDNSNPGKGTDEPKGKTEAYLEAIRKNIEWLK 420
Qy 421 KHDKGNKEDYDLKMRDPIFKQADAYVEKGLDKKEAEAIKRIYSSL 468
Db 421 KHDKGNKEDYDLKMRDPIFKQADAYVEKGLDKKEAEAIKRIYSSL 468
RESULT 4
ABG95925
ID ABG95925 standard; Protein; 468 AA.
XX ABG95925;
AC
XX 10-DEC-2002 (first entry)
XX Human secreted/transmembrane protein PRO5990.
XX Human; secreted protein; transmembrane protein; antirheumatic;
KW antirheumatic; osteopathic; sports-related joint problem;
KW articular cartilage defect; osteoarthritis; rheumatoid arthritis.

XX Homo sapiens.
OS US2002119130-A1.
PN 29-AUG-2002.
XX 06-DEC-2001; 2001US-0006867.
XX 29-OCT-1997; 97US-063435P.
PR 29-OCT-1997; 97US-064215P.
PR 22-APR-1998; 98US-082797P.
PR 29-APR-1998; 98US-083495P.
PR 15-MAY-1998; 98US-085579P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 11-JUN-1998; 98US-088863P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089514P.
PR 16-SEP-1998; 98WO-US19330.
PR 08-MAR-1999; 98WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21194.
PR 22-DEC-1999; 99WO-US30720.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 30-MAR-2000; 2000WO-US08439.
PR 22-MAY-2000; 2000WO-US14042.
PR 02-JUN-2000; 2000WO-US15264.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 01-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32378.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
XX (GETH) GENENTECH INC.
XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI,
XX WPI; 2002-731348/79.
DR N-PSDB; ABS74452.
XX New isolated secreted and transmembrane PRO polypeptide useful for
PT modulating biological activity of a cell, or for treating
PT sports-related joint problems, osteoarthritis or rheumatoid arthritis
XX Claim 20; Fig 150; 399pp; English.
XX The invention relates to an isolated secreted and transmembrane PRO
CC polypeptide having 80 % sequence identity to a sequence appearing
CC as ABG95851-ABG95934 or their associated signal peptide, or a sequence of
CC an extracellular domain of the proteins with their associated signal
CC peptide or lacking its associated signal peptide. Also included are
CC the nucleic acids encoding the proteins, vectors, host cells,
CC fusion proteins and antibodies which specifically bind to the proteins.
CC The proteins are useful for detecting a polypeptide designated as A, B, C
CC or D in a sample suspected of containing an A, B, C or D polypeptide,
CC by contacting the sample with a polypeptide designated as E, F, G, H or
CC I (or vice versa) and determining the formation of a A/E, B/F, G/H or
CC or D/I polypeptide conjugate in the sample, where the formation of the
CC conjugate is indicative of the presence of an A, B, C or D polypeptide
CC in the sample, where A is a PRO10272 polypeptide, B is a PRO20110
CC polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,
CC E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040

CC polypeptide, H is a PRO20233 polypeptide and I is a PRO1890
 CC polypeptide. The sample comprises a cell suspected of expressing the A,
 CC B, C or D polypeptide. The A, B, C, D, E, F, G, H or I polypeptide is labeled with
 CC a detectable label or is attached to a solid support. The proteins are
 CC useful for linking a bioactive molecule to a cell expressing a
 CC polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive
 CC molecule is a toxin, a radiolabel or an antibody. The bioactive molecule
 CC causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies
 CC against them are useful for modulating a biological activity of a cell
 CC expressing a polypeptide designated as A, B, C or D or E, F, G, H, or
 CC I. The cell is killed. The proteins are useful for identifying
 CC agonists or antagonists for the preparation of a medicament useful in
 CC the treatment of a condition which is responsive to the proteins, as
 CC molecular weight markers for protein electrophoresis purposes, and as
 CC therapeutic agents for treating sports-related joint problems,
 CC articular cartilage defects, osteoarthritis or rheumatoid arthritis.
 CC Nucleic acids encoding the proteins are useful as hybridisation probes,
 CC in chromosome and gene mapping, in the generation of anti-sense RNA and
 CC DNA, for the preparation of the proteins, to generate transgenic or
 CC knockout animals which are useful in the development and screening of
 CC therapeutic useful reagents, for chromosome identification, and in gene
 CC therapy. The antibody is useful as a therapeutic agent, in a diagnostic
 CC assay and for affinity purification of the protein from recombinant
 CC cell culture natural sources. The present sequence represents a novel
 CC secreted or transmembrane protein of the invention.

XX Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 23; Length 468;
 Best Local Similarity 100.0%; Pred. No. 4.1e-146;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGFLGTGTLVILVLPQAFKPGSGQSKLHNRLSALRPLNEQIAEAEEDKIKTYPP 60
 DB 1 MGFLGTGTLVILVLPQAFKPGSGQSKLHNRLSALRPLNEQIAEAEEDKIKTYPP 60

QY 61 ENKPGQSNYSFVDNLNLKATEKEKIEKQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
 DB 61 ENKPGQSNYSFVDNLNLKATEKEKIEKQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120

QY 121 DSTKSGLDHKKQDDPDGHLQDGTPLTAEDIVHKIAARIYEENDRAVDKIVSKLNLGL 180
 DB 121 DSTKSGLDHKKQDDPDGHLQDGTPLTAEDIVHKIAARIYEENDRAVDKIVSKLNLGL 180

QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTEQAGKIPKVTMPMAAIODGL 240
 DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTEQAGKIPKVTMPMAAIODGL 240

QY 241 AKGENDETQSVNTLTITNGLRTKTYSDNFPEELQYFNFVALLKSIDSEKEKEKETLI 300
 DB 241 AKGENDETQSVNTLTITNGLRTKTYSDNFPEELQYFNFVALLKSIDSEKEKEKETLI 300

QY 301 TIMKTLIDFVKMWKYGITISPEEGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360
 DB 301 TIMKTLIDFVKMWKYGITISPEEGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360

QY 361 KSHETDSTKEAAKWEKEYGSLKSDTKDDNSNPGCKTDEPKGTEAYLEAIRKNIEWLK 420
 DB 361 KSHETDSTKEAAKWEKEYGSLKSDTKDDNSNPGCKTDEPKGTEAYLEAIRKNIEWLK 420

QY 421 KHKKGKKNEDYDLSKMRDPIKQADAYVEKGILDKKEAEAIKRIYSSL 468
 DB 421 KHKKGKKNEDYDLSKMRDPIKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 5

ABU69097
 ID ABU69097 standard; Protein; 468 AA.

XX AC ABU69097;

XX DT 02-JUN-2003 (first entry)

XX

Human PRO polypeptide #18.

Human; PRO polypeptide; secreted and transmembrane protein; cancer;
 non-insulin dependent diabetes mellitus; septic shock; stroke;
 rheumatoid arthritis; graft-versus-host disease; cardiac ischaemia;
 psoriasis; inflammatory bowel disease; asthma; antidiabetic;
 cycostatic; immunosuppressive; antirheumatic; antiarthritis;
 cerebroprotective; vasotropic; antipsoriatic; antiinflammatory;
 antiasthmatic.

Homo sapiens.

US2003008348-A1.

09-JAN-2003.

26-DEC-2001; 2001US-0035855.

14-MAY-1999; 99WO-US10733.

02-DEC-1999; 99WO-US28551.

22-DEC-1999; 99WO-US30720.

01-MAR-2000; 2000WO-US05601.

02-MAR-2000; 2000WO-US05841.

22-MAY-2000; 2000WO-US14042.

02-JUN-2000; 2000WO-US15264.

23-AUG-2000; 2000WO-US23522.

24-AUG-2000; 2000WO-US23328.

01-DEC-2000; 2000WO-US32678.

01-JUN-2001; 2001WO-US17800.

20-JUN-2001; 2001WO-US19692.

29-JUN-2001; 2001WO-US21066.

09-JUL-2001; 2001WO-US21735.

15-MAY-1998; 98US-085579P.

15-DEC-1998; 98US-112514P.

22-DEC-1998; 98US-113300P.

23-DEC-1998; 98US-113430P.

23-DEC-1998; 98US-113605P.

22-JAN-1999; 99US-116843P.

23-MAR-1999; 99US-125774P.

23-MAR-1999; 99US-125778P.

24-MAR-1999; 99US-125826P.

31-MAR-1999; 99US-127035P.

05-APR-1999; 99US-127106P.

13-APR-1999; 99US-129122P.

21-APR-1999; 99US-130359P.

25-MAY-1999; 99US-135750P.

08-JUN-1999; 99US-138168P.

20-JUL-1999; 99US-144791P.

03-AUG-1999; 99US-146970P.

29-OCT-1999; 99US-162506P.

16-AUG-2001; 2001US-0931836.

(GETH) GENENTECH INC.

Deenoyers LA, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;

Stewart T, Watanabe CK, Wood WI, Zhang Z;

WPI; 2003-341326/32.

N-PSDB; ACA06116.

New PRO polypeptides and nucleic acid molecules, useful for diagnosing
 or treating diabetes mellitus, cancers, septic shock, inflammatory
 bowel disease or asthma, or in gene therapy, chromosome identification.
 or tissue typing

Claim 12; Fig 36; 196pp; English.

The present invention relates to the isolation of novel human PRO
 polypeptides, and the polynucleotide sequences encoding them. The
 PRO polypeptides are secreted and transmembrane proteins. The PRO
 polypeptides and polynucleotides are useful in diagnosing or
 treating non-insulin dependent diabetes mellitus, cancers, septic
 shock, rheumatoid arthritis, graft-versus-host disease, stroke, cardiac

CC ischaemia, psoriasis, inflammatory bowel disease or asthma. The PRO
 CC polynucleotide sequences may be used as hybridisation probes in
 CC chromosome and gene mapping, or in generating antisense RNA and DNA.
 CC They are also useful in preparing PRO polypeptides, in assays to
 CC identify other proteins or molecules involved in binding reaction, to
 CC generate transgenic animals or knockout animals, which in turn are
 CC useful in the development and screening of therapeutically useful
 CC reagents, for chromosome identification, and tissue typing. The PRO
 CC polypeptides and nucleic acid molecules are also useful in gene
 CC therapy, and as molecular weight markers for protein electrophoresis
 CC purposes. Anti-PRO antibodies may be used in diagnostic assays for
 CC PRO polypeptides, or for the affinity purification of the polypeptides
 CC from recombinant cell culture or natural sources. ABU69080-ABU69102
 CC represent the human PRO polypeptides of the invention.
 XX Sequence 468 AA;
 SQ

Query Match 100.0%; Score 2399; DB 24; Length 468;
 Best Local Similarity 100.0%; Pred. No. 4.1e-146;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGWIILVLPQAFKPGQSDKSLHNLRLSAERPLNEQIAEAEEDKIKTYPP 60
 Db 1 MGFLGTGWIILVLPQAFKPGQSDKSLHNLRLSAERPLNEQIAEAEEDKIKTYPP 60
 Qy 61 ENKPGQSNYSFVDNLNLKAI TEKEKEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
 Db 61 ENKPGQSNYSFVDNLNLKAI TEKEKEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
 Qy 121 DSTKSGLDHKKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
 Db 121 DSTKSGLDHKKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
 Qy 181 ITESQAHTELEVAEVLQKLISKANNVEEDPNKPTSWTENQAGKPEKVTWAAIODGL 240
 Db 181 ITESQAHTELEVAEVLQKLISKANNVEEDPNKPTSWTENQAGKPEKVTWAAIODGL 240
 Qy 241 AKGENDETNSNTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
 Db 241 AKGENDETNSNTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
 Qy 301 TIMKTLIDFVKMVKYGTISPPEGVSVLENLDEMIALQTKNLEKATNDNISKLPAPSE 360
 Db 301 TIMKTLIDFVKMVKYGTISPPEGVSVLENLDEMIALQTKNLEKATNDNISKLPAPSE 360
 Qy 361 KSHETDSTKEEAAKMEKYGSLKSDTKDONSPPGGKTDEPKGTAYLEAIRKNIEWLK 420
 Db 361 KSHETDSTKEEAAKMEKYGSLKSDTKDONSPPGGKTDEPKGTAYLEAIRKNIEWLK 420
 Qy 421 KHKKGKNDYDLKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
 Db 421 KHKKGKNDYDLKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 6
 ABU69120
 ID ABU69120 standard; Protein; 468 AA.
 XX
 AC ABU69120;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Human PRO polypeptide #18.
 XX
 KW Human; secreted and transmembrane protein; bone disorder; obesity;
 KW cartilage disorder; sports injury; arthritis; diabetes mellitus;
 KW hypo-insulinaemia; obesity; hyper-insulinaemia; thalassaemia;
 KW haemoglobin-associated disorder; kidney disorder; Berger disease;
 KW mesangial cell function; nephropathy; Schonlein-Henoch purpura;
 KW celiac disease; dermatitis herpetiformis; Crohn's disease;
 KW anorectic; antiarthritic; antidiabetic; antianaemic; nephrotropic;
 XX antiinflammatory.

OS Homo sapiens.
 XX US2003032061-A1.
 XX 13-FEB-2003.
 XX 26-DEC-2001; 2001US-0036214.
 XX 14-MAY-1999; 99WO-US10733.
 XX 02-DEC-1999; 99WO-US28551.
 XX 22-DEC-1999; 99WO-US30720.
 XX 01-MAR-2000; 2000WO-US05601.
 XX 02-MAR-2000; 2000WO-US05841.
 XX 22-MAY-2000; 2000WO-US14042.
 XX 02-JUN-2000; 2000WO-US15264.
 XX 23-AUG-2000; 2000WO-US23522.
 XX 24-AUG-2000; 2000WO-US23328.
 XX 01-DEC-2000; 2000WO-US32678.
 XX 20-DEC-2000; 2000WO-US34956.
 XX 28-FEB-2001; 2001WO-US06520.
 XX 01-JUN-2001; 2001WO-US17800.
 XX 20-JUN-2001; 2001WO-US19692.
 XX 29-JUN-2001; 2001WO-US21066.
 XX 09-JUL-2001; 2001WO-US21735.
 XX 15-MAY-1998; 98US-085579P.
 XX 15-DEC-1998; 98US-112514P.
 XX 22-DEC-1998; 98US-113100P.
 XX 23-DEC-1998; 98US-113430P.
 XX 23-DEC-1998; 98US-113605P.
 XX 23-DEC-1998; 98US-113621P.
 XX 12-JAN-1999; 99US-115552P.
 XX 22-JAN-1999; 99US-116843P.
 XX 23-MAR-1999; 99US-125774P.
 XX 23-MAR-1999; 99US-125778P.
 XX 24-MAR-1999; 99US-125826P.
 XX 31-MAR-1999; 99US-127035P.
 XX 05-APR-1999; 99US-127706P.
 XX 13-APR-1999; 99US-129122P.
 XX 21-APR-1999; 99US-130359P.
 XX 27-APR-1999; 99US-131270P.
 XX 27-APR-1999; 99US-131272P.
 XX 27-APR-1999; 99US-131291P.
 XX 04-MAY-1999; 99US-132371P.
 XX 04-MAY-1999; 99US-132379P.
 XX 04-MAY-1999; 99US-132383P.
 XX 25-MAY-1999; 99US-135750P.
 XX 08-JUN-1999; 99US-138166P.
 XX 20-JUL-1999; 99US-144791P.
 XX 03-AUG-1999; 99US-146970P.
 XX 29-OCT-1999; 99US-162506P.
 XX 16-AUG-2001; 2001US-0931836.
 XX (GETH) GENENTECH INC.
 XX Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
 XX Stewart TA, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2003-341962/32.
 XX N-PSDB; ACA06173.
 XX Novel isolated PRO polypeptides e.g., PRO4334, PRO1122, PRO1889,
 XX PRO1890, PRO1887, PRO1785, PRO4351, useful for treating sports
 XX injuries, arthritis, diabetes, obesity, hyper- or hypo-insulinaemia -
 XX Claim 12; Fig 36; 194pp; English.
 XX The present invention relates to the isolation of novel human PRO
 XX polypeptides, and the polynucleotide sequences encoding them. The
 XX PRO polypeptides are secreted and transmembrane proteins. The PRO
 XX polypeptides and polynucleotides are useful in diagnosing or
 XX treating various bone and/or cartilage disorders (e.g. sports
 XX injuries, arthritis), various insulin deficient states (e.g. diabetes

CC	mellitus, hypo-insulinaemia), obesity, hyper-insulinaemia,	
CC	haemoglobin-associated disorders (e.g. thalassemias), kidney disorders	
CC	associated with decreased mesangial cell function (e.g. Berger disease),	
CC	or other nephropathies associated with Schonlein-Henoch purpura,	
CC	celliac disease, dermatitis herpetiformis or Crohn's disease. The PRO	
CC	polynucleotide sequences may be used as hybridisation probes in	
CC	chromosome and gene mapping, or in generating antisense RNA and DNA.	
CC	They are also useful in preparing PRO polypeptides, in assays to	
CC	identify other proteins or molecules involved in binding reaction, to	
CC	generate transgenic animals or knockout animals, which in turn are	
CC	useful in the development and screening of therapeutically useful	
CC	reagents, for chromosome identification, and tissue typing. The PRO	
CC	polypeptides and nucleic acid molecules are also useful in gene	
CC	therapy, and as molecular weight markers for protein electrophoresis	
CC	purposes. Anti-PRO antibodies may be used in diagnostic assays for	
CC	PRO polypeptides, or for the affinity purification of the polypeptides	
CC	from recombinant cell culture or natural sources. ABU69103-ABU69125	
CC	represent the human PRO polypeptides of the invention.	
XX		
SQ	Sequence 468 AA;	
	Query Match 100.0%; Score 2399; DB 24; Length 468;	
	Best Local Similarity 100.0%; Pred. No. 4.1e-146;	
	Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MGFLGTGWLVLVLPQAPFKPGSQKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60	
DB	1 MGFLGTGWLVLVLPQAPFKPGSQKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60	
QY	61 ENKPGQSNYSFVDNLNLLKAITEKEKIERQSISSPLDNKLVNVEDVSTNKRKLIDDY 120	
DB	61 ENKPGQSNYSFVDNLNLLKAITEKEKIERQSISSPLDNKLVNVEDVSTNKRKLIDDY 120	
QY	121 DSTKSLDGHKFDQDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDKIVSKLLNLGL 180	
DB	121 DSTKSLDGHKFDQDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDKIVSKLLNLGL 180	
QY	181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTFMAAIQDGL 240	
DB	181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTFMAAIQDGL 240	
QY	241 AKGENDETQVNTLTITNGLERRTKTYSDNFEELQYFNFYALLKSIDSEKEAKEFTLI 300	
DB	241 AKGENDETQVNTLTITNGLERRTKTYSDNFEELQYFNFYALLKSIDSEKEAKEFTLI 300	
QY	301 TIMKTLIDFVKMWVYGTISPEEGVSYLENLDEMIALQTKNKLKNQATDNISKLPAPSE 360	
DB	301 TIMKTLIDFVKMWVYGTISPEEGVSYLENLDEMIALQTKNKLKNQATDNISKLPAPSE 360	
QY	361 KSHEETDSTKEEAAKWEYGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNIEWLK 420	
DB	361 KSHEETDSTKEEAAKWEYGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNIEWLK 420	
QY	421 KHKKGKNGEDYLSKQRFINKQADAYVEKGILDKKEAEAKRIYSSL 468	
DB	421 KHKKGKNGEDYLSKQRFINKQADAYVEKGILDKKEAEAKRIYSSL 468	
RESULT 7		
ABU71580		
ID	ABU71580 standard; Protein; 468 AA.	
XX		
AC	ABU71580;	
XX		
DT	10-JUN-2003 (first entry)	
XX		
DE	Human secreted polypeptide PRO5990.	
XX		
KW	Human; gene therapy; tumour; cancer.	
XX		
OS	Homo sapiens.	
XX		
PN	US2003013855-A1.	

XX	16-JAN-2003.	
PD	03-MAY-2002; 2002US-0063616.	
XX		
PF	30-DEC-1998; 98KR-0062142.	
XX	08-MAR-1999; 99WO-US05028.	
PR	14-MAY-1999; 99WO-US10733.	
PR	30-DEC-1999; 99WO-US31274.	
PR	18-FEB-2000; 2000WO-US04341.	
PR	01-MAR-2000; 2000WO-US05601.	
PR	02-MAR-2000; 2000WO-US05841.	
PR	21-MAR-2000; 2000WO-US07532.	
PR	22-MAY-2000; 2000WO-US14042.	
PR	02-JUN-2000; 2000WO-US15264.	
PR	24-AUG-2000; 2000WO-US23328.	
PR	10-NOV-2000; 2000WO-US30873.	
PR	01-DEC-2000; 2000WO-US32678.	
PR	20-DEC-2000; 2000WO-US34956.	
PR	28-FEB-2001; 2001WO-US06520.	
PR	01-JUN-2001; 2001WO-US17800.	
PR	14-MAY-1999; 99US-0311832.	
PR	25-AUG-1999; 99US-0380137.	
PR	25-AUG-1999; 99US-0380138.	
PR	25-AUG-1999; 99US-0380139.	
PR	15-SEP-1999; 99US-0380142.	
PR	15-SEP-1999; 99US-0397342.	
PR	18-OCT-1999; 99US-0403297.	
PR	12-NOV-1999; 99US-0423844.	
PR	22-AUG-2000; 2000US-0644848.	
PR	18-SEP-2000; 2000US-0646610.	
PR	18-SEP-2000; 2000US-0665350.	
PR	08-NOV-2000; 2000US-0709238.	
PR	20-DEC-2000; 2000US-0747259.	
PR	22-MAR-2001; 2001US-0816744.	
PR	10-MAY-2001; 2001US-0854208.	
PR	10-MAY-2001; 2001US-0854280.	
PR	30-MAY-2001; 2001US-0870574.	
PR	05-JUN-2001; 2001US-0874503.	
PR	29-JUN-2001; 2001US-0869599.	
PR	18-JUL-2001; 2001US-0908827.	
PR	06-DEC-2001; 2001US-0006867.	
XX		
PA	(GETH) GENENTECH INC.	
XX		
XX	Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PU;	
PI	Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;	
XX		
DR	WPI; 2003-330485/31.	
DR	N-PSDB; ACA58884.	
XX		
PT	New isolated antibody specifically binding a PRO polypeptide, useful	
PT	for the preparation of a medicament for treating disorders with the	
PT	aberrant expression or activity of the PRO polypeptide, such as tumor	
XX	conditions and cancer	
PS	Example 17; Page 218-219; 406pp; English.	
XX		
CC	The invention relates to an antibody that binds to a polypeptide with a	
CC	fully defined sequence given in the specification. The methods and	
CC	compositions (containing antibodies that specifically bind a PRO	
CC	polypeptide) of the present invention are useful for the preparation of a	
CC	medicament for the treatment of disorders associated with the aberrant	
CC	expression or activity of the PRO polypeptide, such as tumour conditions	
CC	and cancer. They can also be used to generate transgenic or knockout	
CC	animals useful in the development and screening of therapeutically useful	
CC	reagents. The PRO polypeptides and encoding nucleic acids can be used as	
CC	molecular weight markers for protein electrophoresis, chromosome	
CC	identification and tissue typing. The PRO polypeptides are useful to	
CC	induce angiogenesis e.g wound healing; in the treatment of sports-related	
CC	joint problems, articular cartilage defects, osteoarthritis or rheumatoid	
CC	arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The	
CC	antibodies may be used in various diagnostic, competitive binding and/or	

CC immunoprecipitation assays. The present sequence represents the amino
 XX acid sequence of a PRO polypeptide of the invention.
 SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
 Best Local Similarity 100.0%; Pred. No. 4.1e-146;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGFLGTGTVILVLPQAPPGSGQSKLHNRELSAERPLNEQIAEAEDKIKKTYPP 60
 Db 1 MGFLGTGTVILVLPQAPPGSGQSKLHNRELSAERPLNEQIAEAEDKIKKTYPP 60
 QY 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIERQSISSPLDNKLNVEDVSTNKRKLDDY 120
 Db 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIERQSISSPLDNKLNVEDVSTNKRKLDDY 120
 QY 121 DSTKSGLDHKKFQDDPGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
 Db 121 DSTKSGLDHKKFQDDPGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
 QY 181 ITESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIPEKVTMPMAAIOGDL 240
 Db 181 ITESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIPEKVTMPMAAIOGDL 240
 QY 241 AKGENDETVSNTLTITNGLERTKTYSEDNFEELQYFPNPFYALLKSIDSEKAEKETLI 300
 Db 241 AKGENDETVSNTLTITNGLERTKTYSEDNFEELQYFPNPFYALLKSIDSEKAEKETLI 300
 QY 301 TIMKTLIDFVKMWKYGTISPPEGVSYLENDEMIALQTNKLEKNATDNISKLPAPSE 360
 Db 301 TIMKTLIDFVKMWKYGTISPPEGVSYLENDEMIALQTNKLEKNATDNISKLPAPSE 360
 QY 361 KSHEETDSTKEEAAKMEKEYSLKSDTKDSDNSPGKTDPEKGTAEYLAIRKNIEWLK 420
 Db 361 KSHEETDSTKEEAAKMEKEYSLKSDTKDSDNSPGKTDPEKGTAEYLAIRKNIEWLK 420
 QY 421 KHKKGKKNEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
 Db 421 KHKKGKKNEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 8
 ID ABU72026 standard; Protein; 468 AA.
 XX AC ABU72026;
 XX DT 11-JUN-2003 (first entry)
 DE Novel human secreted and transmembrane protein PRO5990.
 KW Human; secreted and transmembrane polypeptide;
 KW chromosome mapping; gene mapping; transgenic animal; knockout animal;
 KW therapeutic agent screening; chromosome identification; tissue typing;
 KW gene therapy.
 XX OS Homo sapiens.
 XX FN US2003018183-A1.
 XX PD 23-JAN-2003.
 XX PP 01-MAY-2002; 2002US-0063512.
 XX PR 06-DEC-2001; 2001US-0006867.
 XX PA (GETH) GENENTECH INC.
 XX PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
 XX Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
 XX WPI; 2003-330984/31.

DR N-PSDB; ACA60437.
 XX New secreted and transmembrane PRO polypeptides and nucleic acid
 PT molecules encoding the polypeptides, useful in gene therapy or
 PT preparing a medicament for treating a condition that is responsive to
 PT the PRO polypeptide or antibody -
 XX Disclosure; Fig 150; 409pp; English.
 PS
 XX The invention describes novel isolated PRO polypeptides. The PRO
 CC polypeptides or anti-PRO antibodies are useful in preparing a medicament
 CC for treating a condition that is responsive to the PRO polypeptide or
 CC antibody. The PRO nucleotide sequences may be used as hybridisation
 CC probes in chromosome and gene mapping, or in generating antisense RNA
 CC and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,
 CC in assays to identify other proteins or molecules involved in binding
 CC reaction, to generate transgenic animals or knockout animals, which in
 CC turn are useful in the development and screening of therapeutically
 CC useful reagents, for chromosome identification, and tissue typing. The
 CC PRO polypeptides and nucleic acid molecules are also useful in gene
 CC therapy, and as molecular weight markers for protein electrophoresis
 CC purposes. The anti-PRO antibodies may be used in diagnostic assays for
 CC PRO, or for the affinity purification of PRO from recombinant cell
 CC culture or natural sources. This is the amino acid sequence of a novel
 CC human secreted and transmembrane PRO polypeptide.
 XX Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
 Best Local Similarity 100.0%; Pred. No. 4.1e-146;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGFLGTGTVILVLPQAPPGSGQSKLHNRELSAERPLNEQIAEAEDKIKKTYPP 60
 Db 1 MGFLGTGTVILVLPQAPPGSGQSKLHNRELSAERPLNEQIAEAEDKIKKTYPP 60
 QY 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIERQSISSPLDNKLNVEDVSTNKRKLDDY 120
 Db 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIERQSISSPLDNKLNVEDVSTNKRKLDDY 120
 QY 121 DSTKSGLDHKKFQDDPGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
 Db 121 DSTKSGLDHKKFQDDPGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
 QY 181 ITESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIPEKVTMPMAAIOGDL 240
 Db 181 ITESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAGKIPEKVTMPMAAIOGDL 240
 QY 241 AKGENDETVSNTLTITNGLERTKTYSEDNFEELQYFPNPFYALLKSIDSEKAEKETLI 300
 Db 241 AKGENDETVSNTLTITNGLERTKTYSEDNFEELQYFPNPFYALLKSIDSEKAEKETLI 300
 QY 301 TIMKTLIDFVKMWKYGTISPPEGVSYLENDEMIALQTNKLEKNATDNISKLPAPSE 360
 Db 301 TIMKTLIDFVKMWKYGTISPPEGVSYLENDEMIALQTNKLEKNATDNISKLPAPSE 360
 QY 361 KSHEETDSTKEEAAKMEKEYSLKSDTKDSDNSPGKTDPEKGTAEYLAIRKNIEWLK 420
 Db 361 KSHEETDSTKEEAAKMEKEYSLKSDTKDSDNSPGKTDPEKGTAEYLAIRKNIEWLK 420
 QY 421 KHKKGKKNEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
 Db 421 KHKKGKKNEDYDLSKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

RESULT 9
 ID ABU72183 standard; Protein; 468 AA.
 XX AC ABU72183;
 XX DT 13-JUN-2003 (first entry)

DE Human PRO polypeptide #75.
XX Human; PRO polypeptide; secreted and transmembrane protein;
KW anti-PRO antibody; diagnostic assay; gene expression.
XX Homo sapiens.
OS
XX US2003023042-A1.
FN 30-JAN-2003.
XX
XX 01-MAY-2002; 2002US-0063502.
XX
XX 06-DEC-2001; 2001US-0006867.
XX
XX (GETH) GENENTECH INC.
PA
XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX WPI; 2003-331484/31.
DR N-PSDB; ACA63447.
XX
XX Novel monoclonal antibody that binds to secreted and transmembrane
PT polypeptide, useful for detecting and purifying the polypeptide and
PT also for treating conditions responsive to the antibody -
XX
XX Disclosure; Fig 150; 408pp; English.
XX
XX The present invention relates to the isolation of novel human PRO
CC polypeptides, and the polynucleotide sequences encoding them. The
CC PRO polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides and polynucleotides are useful for preparing a
CC medicament useful in the treatment of a condition responsive to
CC anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic
CC assays for PRO, by detecting its expression in specific cells,
CC tissues or serum, and for affinity purification of PRO from
CC recombinant cell culture or natural sources. ABU72109-ABU72192
CC represent the human PRO polypeptides of the invention.
XX
XX Sequence 468 AA;
SQ
Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGWIIVLVLPQAFKPGSGQSKLHNRELSAERPLNEQIAEAEDKIKTYPP 60
DB 1 MGFLGTGWIIVLVLPQAFKPGSGQSKLHNRELSAERPLNEQIAEAEDKIKTYPP 60
QY 61 ENKPGQSNYSFVDNLNLLKATEKEKIERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
DB 61 ENKPGQSNYSFVDNLNLLKATEKEKIERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
QY 121 DSTKSGLDHKEQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
DB 121 DSTKSGLDHKEQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEENPNKPTSWTENQAGKIPKVTMPMAIQDGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEENPNKPTSWTENQAGKIPKVTMPMAIQDGL 240
QY 241 AKGENDETVSNLTLTNGLERTKTYSDNPEELQYFNFYALLKSIDSEKEAKEKETLI 300
DB 241 AKGENDETVSNLTLTNGLERTKTYSDNPEELQYFNFYALLKSIDSEKEAKEKETLI 300
QY 301 TIMKTLDIVFVMVKYGTISPEEGVSYLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
DB 301 TIMKTLDIVFVMVKYGTISPEEGVSYLENLDEMIALQTNKLEKNATDNISKLPAPSE 360
QY 361 KSHETDSTKEAAKMEKESGLKSDTKDNNFSGGKTDEPKGTEYLAIRKNIEWLK 420
DB 361 KSHETDSTKEAAKMEKESGLKSDTKDNNFSGGKTDEPKGTEYLAIRKNIEWLK 420

QY 421 KHKDKGNKEDYDLKSMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
DB 421 KHKDKGNKEDYDLKSMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
RESULT 10
ABU66614
ID ABU66614 standard; Protein; 468 AA.
XX
XX AC ABU66614;
XX
XX DT 23-MAY-2003 (first entry)
XX
XX DE Human PRO polypeptide #45.
XX
XX KW Human; PRO polypeptide; secreted and transmembrane protein;
KW tumour necrosis factor-alpha; TNF-alpha; blood; proliferation;
KW differentiation; chondrocyte; tumour; genetic disorder;
KW cytostatic.
XX
XX OS Homo sapiens.
XX
XX PN US2003036180-A1.
XX
XX PD 20-FEB-2003.
XX
XX PF 09-MAY-2002; 2002US-0143114.
XX
XX PR 31-MAR-1997; 97WO-US05230.
PR 12-JUN-1998; 98WO-US12456.
PR 14-JUL-1998; 98WO-US14552.
PR 28-AUG-1998; 98WO-US17888.
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19093.
PR 14-SEP-1998; 98WO-US19094.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 29-OCT-1998; 98WO-US22991.
PR 29-OCT-1998; 98WO-US22992.
PR 20-NOV-1998; 98WO-US24855.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 99WO-US00106.
PR 08-MAR-1999; 99WO-US05028.
PR 10-MAR-1999; 99WO-US05190.
PR 20-APR-1999; 99WO-US08615.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20411.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 22-DEC-1999; 99WO-US30720.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.

XX OS Homo sapiens.
 XX PN US2003032155-A1.
 XX PD 13-FEB-2003.
 XX PF 03-MAY-2002; 2002US-0137865.
 XX PR 31-MAR-1997; 97WO-US05230.
 PR 12-JUN-1998; 98WO-US12456.
 PR 14-JUL-1998; 98WO-US14552.
 PR 28-AUG-1998; 98WO-US17888.
 PR 10-SEP-1998; 98WO-US18824.
 PR 14-SEP-1998; 98WO-US19093.
 PR 14-SEP-1998; 98WO-US19094.
 PR 14-SEP-1998; 98WO-US19177.
 PR 16-SEP-1998; 98WO-US19130.
 PR 17-SEP-1998; 98WO-US19437.
 PR 07-OCT-1998; 98WO-US21141.
 PR 29-OCT-1998; 98WO-US22991.
 PR 29-OCT-1998; 98WO-US22992.
 PR 20-NOV-1998; 98WO-US24855.
 PR 01-DEC-1998; 98WO-US25108.
 PR 05-JAN-1999; 99WO-US00106.
 PR 08-MAR-1999; 99WO-US05028.
 PR 10-MAR-1999; 99WO-US05190.
 PR 20-APR-1999; 99WO-US08615.
 PR 14-MAY-1999; 99WO-US10733.
 PR 02-JUN-1999; 99WO-US12252.
 PR 01-SEP-1999; 99WO-US20111.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 30-NOV-1999; 99WO-US28409.
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 22-DEC-1999; 99WO-US30720.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 11-FEB-2000; 2000WO-US00376.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05746.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 10-MAR-2000; 2000WO-US06319.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 11-AUG-2000; 2000WO-US22031.

PR 23-AUG-2000; 2000WO-US23522.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 10-NOV-2000; 2000WO-US30873.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 PR 28-FEB-2001; 2001WO-US06520.
 PR 01-MAR-2001; 2001WO-US06666.
 PR 25-MAY-2001; 2001WO-US17092.
 PR 01-JUN-2001; 2001WO-US17800.
 PR 20-JUN-2001; 2001WO-US19692.
 PR 22-JUN-2001; 2001WO-US20116.
 PR 29-JUN-2001; 2001WO-US21086.
 PR 09-JUL-2001; 2001WO-US21735.
 PR 20-DEC-2000; 2000US-074259.
 PR 28-FEB-2001; 2001US-0796498.
 PR 09-MAR-2001; 2001US-0802706.
 PR 14-MAR-2001; 2001US-0808689.
 PR 22-MAR-2001; 2001US-0816744.
 PR 05-APR-2001; 2001US-0828366.
 PR 10-MAY-2001; 2001US-0854208.
 PR 10-MAY-2001; 2001US-0854280.
 PR 18-MAY-2001; 2001US-0860216.
 PR 25-MAY-2001; 2001US-0866028.
 PR 25-MAY-2001; 2001US-0866034.
 PR 01-JUN-2001; 2001US-0872035.
 PR 14-JUN-2001; 2001US-0874503.
 PR 19-JUN-2001; 2001US-0882636.
 PR 21-JUN-2001; 2001US-0886342.
 PR 18-JUL-2001; 2001US-0908827.
 PR 06-AUG-2001; 2001US-0924419.
 PR 09-AUG-2001; 2001US-0927796.
 PR 16-AUG-2001; 2001US-0931836.
 PR 19-DEC-2001; 2001US-0028072.

(GETH) GENENTECH INC.

Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 WPI; 2003-331925/31.
 N-PSDB; ACA04068.

New secreted and transmembrane nucleic acids and polypeptides,
 designated as PRO, useful for treating inflammation, organ failure,
 atherosclerosis, cardiac injury, infertility, birth defects, premature
 aging, AIDS, or cancer

Claim 12; Fig 90; 659pp; English.

The invention relates to an isolated nucleic acid comprising, or which is
 at least 80% identical to, or the full-length coding sequence of, any of
 the 275 nucleotide sequences, encoding the corresponding PRO polypeptide
 (one of 275 secreted or transmembrane proteins). The nucleic acid
 further comprises the full-length coding sequence of the DNA deposited
 under American Type Culture Collection (ATCC) accession number in a list
 given in the specification. Also included are vectors and host
 cells for producing PRO proteins, PRO fusion proteins, anti-PRO
 antibodies, PRO extracellular domains and mature sequences, methods
 of detecting PRO proteins, methods for stimulating the release of
 TNF-alpha (tumour necrosis factor alpha) from human blood,
 and the proliferation of differentiation of chondrocyte cells, the
 proliferation of, or gene expression in pericyte cells, the release or
 proteoglycans from cartilage, proliferation of inner ear utricular
 supporting cells, the proliferation of T-lymphocyte cells, the release
 of a cytokine from peripheral blood mononuclear cells (PBMC), or the
 proliferation of endothelial cells), a method for modulating the uptake
 of glucose or free fatty acid (FFA) by skeletal muscle cells,
 a method for inhibiting the binding of A-peptide to factor VIIA,
 or the differentiation of adipocyte cells, a method for detecting the
 presence of a tumour in a mammal and an oligonucleotide probe derived

CC from any of the nucleotide sequences cited above. The nucleic acids and
CC polypeptides are useful for treating inflammatory diseases, organ
CC failure, atherosclerosis, cardiac injury, infertility, birth defects,
CC premature aging, AIDS (acquired immunodeficiency syndrome), cancer, or
CC diabetic complications. The nucleic acids are useful as hybridisation
CC probes, in chromosome and gene mapping, and in generating antisense RNA
CC or DNA. The polypeptides are useful as pharmaceuticals, diagnostics,
CC biosensors or bioreactors. Both are useful in tissue typing.
XX The present sequence represents a PRO protein of the invention.
XX
SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGFWILVLPQAFKPGSQDSKLNHRLSARPLNEQIAEAEEDKIKTYPP 60
DB 1 MGFLGTGFWILVLPQAFKPGSQDSKLNHRLSARPLNEQIAEAEEDKIKTYPP 60
QY 61 ENKPGQSNYSFVDNLNLKAI TEKEKIEKERSQSRSSPLDNKLNVEDVDSTNKRKLIDDY 120
DB 61 ENKPGQSNYSFVDNLNLKAI TEKEKIEKERSQSRSSPLDNKLNVEDVDSTNKRKLIDDY 120
QY 121 DSTKSGLDHKFQDDPDLGLHLDGTLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
DB 121 DSTKSGLDHKFQDDPDLGLHLDGTLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
QY 181 ITESQAHLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIQDGL 240
DB 181 ITESQAHLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIQDGL 240
QY 241 AKGNDETVSNLTLTNGLERTKTYSDNPEELQYFPNFYALLKSIDSEKEAKEKETLI 300
DB 241 AKGNDETVSNLTLTNGLERTKTYSDNPEELQYFPNFYALLKSIDSEKEAKEKETLI 300
QY 301 TIMKTLIDFVQWVKYGTISPPEGVSVLENDEMIALQTKNLEKNATDNISKLPAPSE 360
DB 301 TIMKTLIDFVQWVKYGTISPPEGVSVLENDEMIALQTKNLEKNATDNISKLPAPSE 360
QY 361 KSHETSTKEAAKMEKVSGLKSDTQDNNPQGGTDEPKGTTEAYLEAIRKNIEWLK 420
DB 361 KSHETSTKEAAKMEKVSGLKSDTQDNNPQGGTDEPKGTTEAYLEAIRKNIEWLK 420
QY 421 KHKDKGNKEDYDLKMRDFINKQADAYVEKGILDEKAEAKRIYSSL 468
DB 421 KHKDKGNKEDYDLKMRDFINKQADAYVEKGILDEKAEAKRIYSSL 468

RESULT 12
ABUS9695
ID ABUS9695 standard; Protein; 468 AA.
XX
XX AC ABUS9695;
XX
XX
DT 13-MAY-2003 (first entry)
XX
XX
DE Novel secreted and transmembrane protein PRO5990.
XX
XX Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;
KW cardiac insufficiency disorder; cancer; tumour; immune response;
KW adrenal cortical capillary endothelial growth; c-fos induction;
KW vascular endothelial growth factor inhibition; VEGF inhibition;
KW endothelial cell growth inhibitor; T-lymphocytes stimulation;
KW retinal neurons cell survival; rod photoreceptor cell survival;
KW retinal disorder; retinitis pigmentosa; kidney disorder;
KW mammalian kidney mesangial cell proliferation; Berger disease;
KW dermatitis; herpeticiformis; Crohn's disease; chondrocyte proliferation;
KW chondrocyte redifferentiation; sports injury; arthritis.
XX
XX Homo sapiens.
XX
PN US2003017563-A1.

XX 23-JAN-2003.
PD
XX 07-MAY-2002; 2002US-0140808.
PF
XX 31-MAR-1997; 97WO-US05230.
PR 12-JUN-1998; 98WO-US12456.
PR 14-JUL-1998; 98WO-US14552.
PR 28-AUG-1998; 98WO-US17888.
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19093.
PR 14-SEP-1998; 98WO-US19094.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 29-OCT-1998; 98WO-US22991.
PR 29-OCT-1998; 98WO-US22992.
PR 20-NOV-1998; 98WO-US24855.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 98WO-US00106.
PR 08-MAR-1999; 98WO-US05028.
PR 10-MAR-1999; 98WO-US05190.
PR 20-APR-1999; 98WO-US08615.
PR 14-MAY-1999; 98WO-US10733.
PR 02-JUN-1999; 98WO-US12252.
PR 01-SEP-1999; 98WO-US20111.
PR 08-SEP-1999; 98WO-US20594.
PR 13-SEP-1999; 98WO-US20944.
PR 15-SEP-1999; 98WO-US21090.
PR 15-SEP-1999; 98WO-US21547.
PR 05-OCT-1999; 98WO-US23089.
PR 29-NOV-1999; 98WO-US28214.
PR 30-NOV-1999; 98WO-US28313.
PR 30-NOV-1999; 98WO-US28409.
PR 01-DEC-1999; 98WO-US28301.
PR 01-DEC-1999; 98WO-US28634.
PR 02-DEC-1999; 98WO-US28551.
PR 02-DEC-1999; 98WO-US28564.
PR 02-DEC-1999; 98WO-US28565.
PR 16-DEC-1999; 98WO-US30095.
PR 20-DEC-1999; 98WO-US30911.
PR 20-DEC-1999; 98WO-US30999.
PR 22-DEC-1999; 98WO-US30720.
PR 30-DEC-1999; 98WO-US31243.
PR 30-DEC-1999; 98WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05746.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.

XX 17-APR-1998; 98US-0062601.
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Corley NC, Patterson C;
XX
XX WPI; 2003-238117/23.
XX N-PSDB; ABX11115.
XX
XX New human neurosecretory proteins (HUNSP) and polynucleotides, useful
XX for diagnosing, treating or preventing e.g. epilepsy, Alzheimer's
XX disease, schizophrenia, Grave's disease, leukaemia, lymphoma, or breast
XX or brain cancer -
XX
XX Claim 1; Fig 1; 45pp; English.
XX
XX The present invention relates to the isolation of human neurosecretory
XX proteins -1 and -2 (HUNSP-1 and HUNSP-2), and the polynucleotide
XX sequences encoding them. The HUNSP polypeptides, polynucleotides, and
XX agonists are useful for treating a disease or condition associated with
XX decreased expression of functional HUNSP. Antagonists of HUNSP are
XX useful for treating a disease or condition associated with
XX overexpression of functional HUNSP. HUNSP antibodies are useful for
XX diagnosing a condition or disease associated with the expression of
XX HUNSP. The sequences of the invention are particularly useful for
XX diagnosing, treating or preventing neurological disorders
XX (e.g. epilepsy, Huntington's disease, dementia, Alzheimer's disease,
XX multiple sclerosis, Parkinson's disease, schizophrenia, amnesia),
XX endocrine disorders (e.g. hypogonadism, Sheehan's syndrome, diabetes
XX insipidus, Hashimoto's disease, Grave's disease, Cushing's disease),
XX and cancers (e.g. adenocarcinoma, leukaemia, lymphoma, melanoma,
XX myeloma sarcoma, bone cancer, breast cancer, cancer of the pancreas,
XX brain cancer, prostate cancer, skin cancer or thyroid cancer).
XX The present sequence represents HUNSP-1. The cDNA encoding HUNSP-1
XX is isolated from pancreatic islet cell cDNA library (ISLNTOT01) Incyte
XX clone 2379427.
XX
XX Sequence 468 AA;
XX
XX Query Match 100.0%; Score 2399; DB 24; Length 468;
XX Best Local Similarity 100.0%; Pred. No. 4.1e-146;
XX Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MGFLGTGTTWLVLPQAPFKPGGQDLSLHNRSLAERPLNEQIAEAEEDKIKKTYPP 60
Db 1 MGFLGTGTTWLVLPQAPFKPGGQDLSLHNRSLAERPLNEQIAEAEEDKIKKTYPP 60
QY 61 ENKPGQSNYSFVNDLNLKAI TEKEKERQSISSPLDNKLNVEDVDSTKNRKLIDDY 120
Db 61 ENKPGQSNYSFVNDLNLKAI TEKEKERQSISSPLDNKLNVEDVDSTKNRKLIDDY 120
QY 121 DSTKSLDGHKFDQDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
Db 121 DSTKSLDGHKFDQDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPKFTSWENQAGKIPKVTPTMAIQDGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPKFTSWENQAGKIPKVTPTMAIQDGL 240
QY 241 AKGENDETYSNTLTITNGLERTKYSEDNPFELQVFNPFYALLKSIDSEKEKEKTLI 300
Db 241 AKGENDETYSNTLTITNGLERTKYSEDNPFELQVFNPFYALLKSIDSEKEKEKTLI 300
QY 301 TIMKTLIDFVQMKVYGTITSPREGVSYLENLDEMIALQTKNLEKXNATDNISKLPFAPSE 360
Db 301 TIMKTLIDFVQMKVYGTITSPREGVSYLENLDEMIALQTKNLEKXNATDNISKLPFAPSE 360
QY 361 KSHHEETDSTKEAAKWEYGLKSTKDDNSNPGKTDPEKGTKEAYLEAIRKNIEWLK 420
Db 361 KSHHEETDSTKEAAKWEYGLKSTKDDNSNPGKTDPEKGTKEAYLEAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468

Db 421 KHDKGNKEDYDLKMRDFINKQADAYVEKGILDKKEAEAIKRIYSSL 468
RESULT 14
AAO13905
ID AAO13905 standard; Protein; 567 AA.
XX AAO13905;
XX
XX 06-NOV-2001 (first entry)
XX Human polypeptide SEQ ID NO 27797.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX Homo sapiens.
XX WO200164835-A2.
XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX N-PSDB; AAI93836.
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX Claim 20; SEQ ID NO 27797; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 567 AA;
XX
XX Query Match 100.0%; Score 2399; DB 22; Length 567;
XX Best Local Similarity 100.0%; Pred. No. 5.3e-146;
XX Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MGFLGTGTTWLVLPQAPFKPGGQDLSLHNRSLAERPLNEQIAEAEEDKIKKTYPP 60
Db 100 MGFLGTGTTWLVLPQAPFKPGGQDLSLHNRSLAERPLNEQIAEAEEDKIKKTYPP 159
QY 61 ENKPGQSNYSFVNDLNLKAI TEKEKERQSISSPLDNKLNVEDVDSTKNRKLIDDY 120
Db 160 ENKPGQSNYSFVNDLNLKAI TEKEKERQSISSPLDNKLNVEDVDSTKNRKLIDDY 219
QY 121 DSTKSLDGHKFDQDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGL 180

Db 220 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDI VHKAARIYEENDRAVDFKIVSKLLNLGL 279
Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIOQGL 240
Db 280 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIOQGL 339
Qy 241 AKGENDETIVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFVALLKSIDSEKEAKEKETLI 300
Db 340 AKGENDETIVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFVALLKSIDSEKEAKEKETLI 399
Qy 301 TIMKTLDIFVKMVKYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 400 TIMKTLDIFVKMVKYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 459
Qy 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDDNSNFGKTDPEPKGTEAYLEAIRKNIEWLK 420
Db 460 KSHEETDSTKEEAAKMEKEYGSLKSDTKDDNSNFGKTDPEPKGTEAYLEAIRKNIEWLK 519
Qy 421 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
Db 520 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 567

RESULT 15

AAW93910
ID AAW93910 standard; Protein; 468 AA.

XX AC AAW93910;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide, SEQ ID NO: 4061.
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX OS Homo sapiens.
XX PN EP1130094-A2.
XX PD 05-SEP-2001.
XX PF 07-JUL-2000; 2000EP-0114089.
XX PR 08-JUL-1999; 99JP-0194486.
XX PR 11-JAN-2000; 2000JP-0118774.
XX PR 02-MAY-2000; 2000JP-0183765.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otauki T, Koga H;
XX DR WPI; 2001-524255/58.
XX DR N-PSDB; AAK94872.

PT 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation.

PS Claim 8; SEQ ID NO 4061; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX SQ Sequence 468 AA;
Query Match 99.7%; Score 2392; DB 22; Length 468;
Best Local Similarity 99.8%; Pred. No. 1.2e-145;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MGFLGTCTWLVLVLPLOAFPKPGGSDKSLHNRRLSAERPLNEQIAEAEEDKIKKTYPP 60
Db 1 MGFLGTCTWLVLVLPLOAFPKPGGSDKSLHNRRLSAERPLNEQIAEAEEDKIKKTYPP 60
Qy 61 ENKPGQSNYSFVNDNLNLLKAITKEKTEKERQSRSSPDLNKLNVEDVDSTKNRKLDDY 120
Db 61 ENKPGQSNYSFVNDNLNLLKAITKEKTEKERQSRSSPDLNKLNVEDVDSTKNRKLDDY 120
Qy 121 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDI VHKAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDI VHKAARIYEENDRAVDFKIVSKLLNLGL 180
Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIOQGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIOQGL 240
Qy 241 AKGENDETIVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFVALLKSIDSEKEAKEKETLI 300
Db 241 AKGENDETIVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFVALLKSIDSEKEAKEKETLI 300
Qy 301 TIMKTLDIFVKMVKYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 301 TIMKTLDIFVKMVKYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360
Qy 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDDNSNFGKTDPEPKGTEAYLEAIRKNIEWLK 420
Db 361 KSHEETDSTKEEAAKMEKEYGSLKSDTKDDNSNFGKTDPEPKGTEAYLEAIRKNIEWLK 420
Qy 421 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
Db 421 KHDKKGKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468

Search completed: January 2, 2004, 11:01:00
Job time : 48 secs

THIS PAGE BLANK (003870)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:32 ; Search time 21 Seconds
(without alignments)
942.928 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTLVLVLPIQAF.....EKGILDKERAEAKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCITUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179	7.5	1786	3	US-08-973-462-8
2	172.5	7.2	1164	3	US-08-923-992A-2
3	169	7.0	1098	3	US-08-923-992A-8
4	168	7.0	1164	3	US-08-923-992A-10
5	167.5	7.0	957	4	US-09-914-259-16
6	166	6.9	1104	3	US-08-923-992A-6
7	166	6.9	1128	3	US-08-923-992A-6
8	166	6.9	2662	4	US-09-595-684B-31
9	163.5	6.8	956	4	US-09-914-259-17
10	162	6.8	8991	4	US-08-714-741-32
11	159.5	6.6	816	2	US-08-533-306A-6
12	159.5	6.6	816	2	US-08-742-923A-6
13	158.5	6.6	885	2	US-08-533-306A-4
14	158.5	6.6	885	2	US-08-742-923A-4
15	156.5	6.5	1087	4	US-09-914-259-12
16	156	6.5	1354	3	US-08-685-871-2
17	154.5	6.4	1231	4	US-08-714-741-41
18	153.5	6.4	984	1	US-08-242-932-2
19	153.5	6.4	984	1	US-08-714-481-2
20	153.5	6.4	984	5	PCT-US95-06111-2
21	153	6.4	1588	4	PCT-US93-07261-11
22	153	6.4	1663	5	PCT-US93-07261-16
23	153	6.4	2285	4	US-09-308-375-2
24	150.5	6.3	1010	4	US-09-134-001C-5178
25	147	6.1	534	4	US-09-103-684A-2
26	146.5	6.1	1388	2	US-08-685-576-1
27	146.5	6.1	1388	2	US-08-685-576-4

28	145.5	6.1	652	4	US-08-559-896B-2	Sequence 2, Appli
29	145	6.0	588	4	US-08-714-741-42	Sequence 42, Appl
30	144.5	6.0	343	3	US-08-937-271-17	Sequence 17, Appl
31	144	6.0	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
32	143.5	6.0	864	4	US-08-714-741-40	Sequence 40, Appl
33	143	6.0	3878	4	US-09-914-259-11	Sequence 11, Appl
34	142	5.9	967	4	US-09-914-259-21	Sequence 21, Appl
35	142	5.9	1032	4	US-09-914-259-26	Sequence 26, Appl
36	141.5	5.9	1146	4	US-08-914-999-6	Sequence 6, Appli
37	141	5.9	1507	3	US-08-929-329-5	Sequence 5, Appli
38	139.5	5.8	963	4	US-09-914-259-20	Sequence 20, Appl
39	139	5.8	1939	4	US-09-310-187A-1	Sequence 1, Appli
40	139	5.8	3111	2	US-08-460-309-4	Sequence 4, Appli
41	139	5.8	3111	2	US-08-125-077-4	Sequence 4, Appli
42	138	5.8	630	3	US-08-973-462-9	Sequence 9, Appli
43	138	5.8	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
44	138	5.8	1234	4	US-09-592-054-8	Sequence 8, Appli
45	137.5	5.7	586	2	US-08-630-822A-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1

US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match	7.5%	Score 179;	DB 3;	Length 1786;
Best Local Similarity	21.5%	Pred. No. 0.00012;		
Matches	118;	Conservative	96;	Mismatches 197; Indels 138; Gaps 25;
Qy	27	QDKSLHNRLSAE--RPLNEQIAEAEEDKIKYPPENKPGQSN-----YSFVDNL--	75	
Db	957	EEAVESNENVAENLEKLNFTVTLVDKVEETVEISGESLENNEMDKAFSEIFDNVKG	1016	
Qy	76	---NLK-----AITEKEKIEKEROISRSSPLDNKLN-----VEDVDSTK--	112	
Db	1017	IQENLLTGMFRSTSTSIQSEKVDL--NENNVSSILDNIENKKEGLLNKLENISSTEGV	1075	
Qy	113	-----NRKLDDYD-----STKGLDHKFDPPDGLHQLDGTPLTAEDI-	151	
Db	1076	QETVTEHVEQNYYVDVDPAMKQDFLGILNEAGLKEMFNFLEL-VFKSESVDITVEEIK	1134	
Qy	152	-----VHKIARIYEENDRAVDFKVLKLNGLLTESQAHLTLEDEVAEVLQKLSKE	204	
Db	1135	DEPVQKEVEKETVSIIEEMEENIVD-----VLEEKEKDLTKMDAVEESEIS	1183	
Qy	205	ANNYEEDPNKPTSTWENOAGKIPEKTPMAAIQDLAKGENDETSTVNTLTITNGLERRTK	264	
Db	1184	SDSKEE-----TES-----IKDKKDVSLVVEEVDNDMDSEVKLELKNWEEELMK	1231	
Qy	265	TYSEDN-----FEELQYFPNFVA--LLKSIDSKEAK-----EKETLITIMK	304	

```

Db 1232 DAVEINDITSKLIBETQELNEVEADLIKDMKLEKLEKALSDSKSKEIIDAADDTLEKVE 1291
QY 305 -----TLIDFVKWVKYGTISPE--EGVSYLENLDEMIALQTK--NKLKRNADTNISKL 354
Db 1292 BEHDIITLIDDEV---VELKOVDEEKIEKVSOLKLEEDILKEVKEIKELSEILEDDYKEL 1348
QY 355 FPAPEKSHETDSTKEEAAKMEKEYGLKSDTKDDNSNPGGKTDEPKGKTEAYLEAIRK 414
Db 1349 KTIETDILEEKKEIEKOHFEKFEFEAEIKOLEADILKEVSSLEVEBEKKLEEVHE-LKE 1407
QY 415 NIEWLKHDK--KNGKEDYLSKWRDFNKQADAY---VEKGILDK-----EEA 458
Db 1408 EVEHIIISGDAHIKGLEED-DLEEVDDLKSGSLDMLKGMELGDMDKESLEDVTTKLGERV 1466
QY 459 EAIKRIYSS 467
Db 1467 ESLKDVLS 1475

RESULT 2
US-08-923-992A-2
; Sequence 2, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IGA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-2

Query Match 7.28; Score 172.5; DB 3; Length 1164;
Best Local Similarity 21.68; Pred. No. 0.0002;
Matches 124; Conservative 77; Mismatches 189; Indels 185; Gaps 25;

QY 36 LSAREPLNE-----QIAEABEDKIKTTPPENKPGQSNYSVDNLLNLLKAIT-EKEK 86
Db 149 LELENQFNENRLLHIIKHQHEVEKDKAK-----QQKTLKQSDTKVDLSNIDKELNHQSKQ 204
QY 87 IEK-ERQSTRSSPLDNKL-----NVEDVDSTKRNKLIDDYDSTKSGLDHK 130

```

```

Db 205 VEKMAEQKGITNEKDSMLKKIEDIRKQAQQADKEDAEVKVREBELKGLFSSTKAGLDQE 264
QY 131 FODDPDGLHQLDGTPLTAEDIVHKI-----AARIYEENDRAVFDKIVSKLLNLUG 179
Db 265 IQE-----HVKKET--SSEENTQKVDHEYANSLQNAQKSLLEELDKATTTNEQATQVKNQF 317
QY 180 L-----ITESQAHTLEDEVAEV-----LOKLISKE---ANNYEE 210
Db 318 LENAQKLKEIOPLIKETNVKLYKAMSESLQVEKELKHNSANLEDLVAKSKEIVREYEG 377
QY 211 DPNKPTSWTE-----NOAGKIPEKVTPTMAAIOGDLAKGENDETVPN 251
Db 378 KLNQSKNLPQLQLEEEAHSKLQVVEDFRKFKTSEQVTPKRVKRDLAANNQ---Q 434
QY 252 TLTLTNGLERTKTYSNEDNFEELQYFFNFYALLXSIDSEKAKEKETITINKTILDFVK 311
Db 435 KIETVSPENITVYEGED-----VKFTVTAKS-DS-----KTLDDFS 471
QY 312 MVKYG-TISPEGVSYLENLD-----EMIALQTK---NKLEKNATD 349
Db 472 LTKYNSVSDRISTNYKTNTDNHKAIEITIKNLKLNESQVTLKAKDDSGNVVETFTI 531
QY 350 NISKLFPAPSEKSHETDSTKEEAA-----KMEKYGSLKD 385
Db 532 TVQKKEEKQVPTPEQKDSKTEKVPQEPKSNQKQLQELIKSAQOQLEKLEKAIKELME 591
QY 386 STKDDNSNPGGKTDEPKGKTEAYLEAIRKNTLEWLK-----HDKKNGKEDYLSKWRDFIN 441
Db 592 Q-PEIPSNP--EYGIQKSIWESQKEPQEAITSFKIIGDSSSKYYTEHYFNKYKSDPMN 648
QY 442 KQADAYVEKGILDKKEAE-----AIKRIYSS 467
Db 649 YQLHAQME--MLTRKVVQVMKYPDPAIEIKIFES 681

RESULT 3
US-08-923-992A-8
; Sequence 8, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IGA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-8

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-8

Query Match          7.0%; Score 169; DB 3; Length 1098;
Best Local Similarity 19.3%; Pred. No. 0.00034;
Matches 104; Conservative 80; Mismatches 182; Indels 174; Gaps 19;

QY 25 GSQKSLHNRELSAERPLNEQIAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLKAI 81
Db 74 GKREKQLOQWKNNLKNVDNLTLSHEQKNBPKTKIDETNDSALLELENFQNETNRLHI 133
QY 82 TEKEKIERQSIRSPIL---DNKLNVEDVDSTKNR-----KLIDQ--- 119
Db 134 KQHEBEVKDKAKAQKQKTLKQSDTKVDLSNIDKELNHQKSQEAGITNEDKDSMLKKIEDIR 193
QY 120 -----YDSTKSGLDHFKQDDPDGHLQDGLDGTPLTAED 150
Db 194 KQAQPKKEDAEVKKVREELGKLFSSSTKAGLDQEIQE-----HVKKET--TSEE 277
QY 156 -----AARIYEENDRAVDFKIVSKLNLGL-----ITESQA 186
Db 247 EHYANSLQNLAKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIKETNVKLYKAMS 306
QY 187 HTLEDEVAEV-----LQKLISKE---ANNVEEDPNKPTSWTE----- 220
Db 307 ESLEQVEKELHNSANLELDVAKSKEIVREYEGKLNQSKNLPKQLEBAHAKLQKV 366
QY 221 ---NQAGKIPKVTMPMAAIOGLAKGENDETNSNTLTNGLERRTKTYSDNFEELQYF 277
Db 367 EDRFRKPKTSQVTPKPKRLDLAANNQ---QKIELTVSPENITVYEGD----- 415
QY 278 PNFVALLKSISEKEAKEKETLIITMTLIDPFVGMVKYG-TISPEEGSVYLENLD----- 332
Db 416 VKFTVTKAS-DS-----KTLTDFSDLLTKYNPSVSDRISTNTKNTDNHKI 460
QY 333 -----EMIALQTK---NKLEKATDNISKLPFPAPSEKSHSEETDSTKEAAK 375
Db 461 AEITIKNLKINESQTVTLKADSGNVVEKFTTITVQKKEKQVTPKPEQKDSKTEKVP 520
QY 376 MEKYGSLKSDTKDSDNSNPGGKTPKGTBAYLEAIRKNIEWLKHGDKGNKEDYDLISK 435
Db 521 QEPK-----SNDKNQLQELIKSAQOELEKLEKAIKELMEQPEIFSNPEYGIQK 568

```

RESULT 4

US-08-923-992A-10
 ; Sequence 10, Application US/08923992A
 ; Patent No. 6280738

GENERAL INFORMATION:

APPLICANT: Tai, Joseph Y.
 APPLICANT: Blake, Milan S.
 TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
 TITLE OF INVENTION: Streptococcal Beta Antigens
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, N.W., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923,992A
 FILING DATE: 05-SEP-1997

```

; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-10

```

Query Match 7.0%; Score 168; DB 3; Length 1164;
 Best Local Similarity 19.0%; Pred. No. 0.00044;
 Matches 121; Conservative 85; Mismatches 177; Indels 254; Gaps 25;

```

QY 25 GSQKSLHNRELSAERPLNEQIAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLKAI 81
Db 105 GKREKQLOQWKNNLKNVDNLTLSHEQKNBPKTKIDETNDSALLELENFQNETNRLHI 164
QY 82 TEKEKIERQSIRSPIL---DNKLNVEDVDSTKNR-----K 115
Db 165 KQHEBEVKDKAKAQKQKTLKQSDTKVDLSNIDKELNHQKSQVETMAEQIGITNEDKDSMLK 224
QY 116 LIDQ-----YDSTKSGLDHFKQDDPDGHLQDGLDGTPLTAED 150
Db 225 KIEDIRKQAQADKEDAEVKKVREELGKLPSTKAGLDQEIQE-----HVKKET--TSEE 277
QY 151 IVHKI-----AARIYEENDRAVDFKIVSKLNL-----LGLITE----- 183
Db 278 NTQKVDHYPNSLQNLAKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIKETNVK 337
QY 184 -----SQAHTLEDEVAEVQKLISKE---ANNVEEDPNKPTSWTE--- 220
Db 338 LYKAMSSLEQVEKQKLNHNSQAN-LEDLVAK-----SKEIVREYEGKLNQSKNLPKQ 390
QY 221 -----NQAGKIPKVTMPMAAIOGLAKGENDE-----TVS-NTLTLTNG 258
Db 391 LEEBAHKLQVVEDFRKPKTSQVTPKPKRDLAANNQKIELTVSPENITVYEG 450
QY 259 LERRTKTYSBDFEELQYFPNFYALLKSIIDSEKEAKEKETLIITMTLIDPFVGMVKYG- 317
Db 451 -----EDLK-----FTLTAKSDS-----KTLTDFSDLLTKYNP 478
QY 318 TISPEEGSVYLENLD-----EMIALQTK---NKLEKATDNISKLP 356
Db 479 SVSDRISTNTKNTDNHKNIAEITIKNLKINESQTVTLKADSGNVVQKFTTITVQKKEE 538
QY 357 APSEKSHSEETDSTKEBAKMEKEVGLSKDSDNSNPGGKTPKGTBAYLEAIRKNI 416
Db 539 KQVPKTPKPEQKDSKTEKVPQEPK-----SNDKNQLQELIKSAQOELEKLEKAI 586
QY 417 EWLKHKHDKGNKEDYDLISK-----AIKRIYSS 467
Db 587 KELMEQPEIFSNPEYGIQKSINESQKEPIQEAITSFKKIIGDSSSKYYTTEHFNKYSDF 646
QY 440 INKQADAYVEKGLDKEEAE-----AIKRIYSS 467
Db 647 MNYQLHAQME--MLTRKVVQVINKYPONABIKKIFES 681

```

RESULT 5

US-09-914-259-16
 ; Sequence 16, Application US/09914259
 ; Patent No. 6495336
 ; GENERAL INFORMATION:

US-08-923-992A-6
 ; Sequence 6, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,992A
 ; FILING DATE: 05-SEP-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/024,707
 ; FILING DATE: 06-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1128 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-923-992A-6

Query Match 6.9%; Score 166; DB 3; Length 1128;
 Best Local Similarity 19.3%; Pred. No. 0.00059;
 Matches 119; Conservative 86; Mismatches 200; Indels 210; Gaps 23;
 QY 25 GSQKSLHNRELSAERPLN---EQIAEAEEDKIKKTYPPENKPGQSNYSP---VDNLNLKAI 81
 DB 69 GKREKQLOQWNNLNDVNTILSHEQKNEFKTKIDETNDSALLELENQFNETNRLHI 128
 QY 82 TEKEKIERQSISSPL---DNKLNVEDVDSTKNR-----K 115
 DB 129 KQHEVEKDKAKQKQTLKQSDTKVLSNIDKELNHQKSPVEKMAEPKGIETNEKDSMLK 188
 QY 116 LIDD-----YDSTKSLDHPKQDDPDGHLQDGLTPLTAED 150
 DB 189 KIEDIRKQAQADKEDAEVKVREELKGLFSSTRAGLDQE-----IHEVHKETSSEE 241
 QY 151 IVHKI-----AARIYEENDRAVDFKIVSKLNLGL----- 180
 DB 242 NTQKVDEHYANSLONLAQKSLLEEDKATTNQEQATQVKNQFLENAAQKLKEMQPLIKETNVK 301
 QY 181 ITESQAHTLEDEVAEVL---LQKLISKE---ANNYEEDPNKPTSWTE----- 220
 DB 302 LYKAMESLEQVEKELKNSANLEEDLVAKSKEIVREVEGKLNQSKNLPKQLKEEAMHS 361
 QY 221 -----NOAGKIPKVTMAAIQDGLAKGENDETVSNLTTLTNGLERRTKTSYDNF 271
 DB 362 KKKQVVEDFRKKFTSEQVTPKRVKRDLANENNQ---QKIELTVSPENITVVEGED-- 416
 QY 272 BELQYFPNFYALLKSIDSEKAKEKETLITMTKLIDFVKRMVKYG-TISPEGVSYLEN 330

DB 417 -----VKFTVTAKS-DS-----KTTLDPSDLLTKYNPSVSDRISTNYKTN 455
 QY 331 LD-----EMIALQTK-----NKLEKNATDNISKLPAPSEKSHETDST 369
 DB 456 TDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVVEKFTTITVQKKEEKQVPKTEQKDSK 515
 QY 370 KEENA-----KWEKEYGSLKSDTKDDNSNPGKTDPEPKGT 405
 DB 516 TEEKVPOEPKSNQNLQELIKSAQOELEKLEKAIKELMEQ-PEIPSNP--EYGIQKSIW 572
 QY 406 EAYLEARKNIEWLK-----HDKKGNKEDYDLSQMRDFINKQADAVYVEKGLDKEEAE-- 459
 DB 573 ESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDPMYQLHAQME--MLTRKVVQYM 630
 QY 460 -----AIKRIYSS 467
 DB 631 NKYPDNAEIKKIPES 645
 RESULT 8
 US-09-595-684B-31
 ; Sequence 31, Application US/09595684B
 ; Patent No. 6544766
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Ohashi, Cara
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Vaisberg, Eugeni
 ; APPLICANT: Wood, Kenneth
 ; APPLICANT: Yu, Ming
 ; TITLE OF INVENTION: Human kinesins and methods of producing
 ; FILE REFERENCE: cytop036
 ; CURRENT APPLICATION NUMBER: US/09/595,684B
 ; PRIOR FILING DATE: 2002-06-24
 ; PRIOR APPLICATION NUMBER: 09/295,612
 ; PRIOR FILING DATE: 2000-04-20
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 31
 ; LENGTH: 2662
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-595-684B-31

Query Match 6.9%; Score 166; DB 4; Length 2662;
 Best Local Similarity 21.8%; Pred. No. 0.0019;
 Matches 119; Conservative 103; Mismatches 175; Indels 150; Gaps 28;
 QY 26 SQKSLHNRELSAERPLN---EQIAEAEEDKIKKTYPPENKPGQSNYSPVDNLNLKAIT 82
 DB 1101 AQEK---NHAIKKEGELSRCTDLAEVEEKLKESQQLQEKQQL-----LNVQEEMS 1150
 QY 83 EKKEKIERQSISSPLDNKLNVEDVDSTK---NRKLIIDDYDSTKS-----GLDHK 130
 DB 1151 EMOKKINEIENKLNKELTLEHMETERLELAQKLNENYEVKSTKERKVLKELQKS 1210
 QY 131 FODDPDGL-----HOLDGTPL-TAEDI---VH-KIAARIYEENDRAVDFKIVSKLNLGL 180
 DB 1211 FETERDHLRGIYIEIEATGLQTEELKIAHILKEHOETIDELRRSVSEK-TAQIINTQD 1269
 QY 181 ITESQAHTLEDEVAEVL---QKLIS--KEANNYEEDPNKPTSWTENQAGKIPKVTMAA 235
 DB 1270 LEKS--HTKLQEEIPLVHBEQQLPNVKYVSEQTETNNELELLTEQSTTK---DSTTLAR 1324
 QY 236 IQ-----DGLAK-GENDETVENTL-----TLTNGLERRTK 264
 DB 1325 IEMERLRLNEKFOESQBEIKSLTKERDNLTKIKEALEVQDQLKEHIRETLAKIQESQSK 1384
 QY 265 TYSEDNFEEL-----QYFPNFYALLK-----SIDSEK-AKEKE 297
 DB 1385 QEOSLNKNEKDNETTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDEMKSVAKEKD 1444

Qy 238 TLITIMKTLDIFVKMMVYGTISPEGVSYLENLDMIA--LOTNKLK-----KNATD 349
Db 1445 DLQRLQEVL-----QSESDQLENIEIVAKHLETEEBELKVAHCLKEQEE 1490
Qy 350 NISKLFPAPSKSHEETSTKEAAKMEKYGSLKSDKSTDDNSNPGGKTDBPKGKTE--- 406
Db 1491 TINELRVNLSEK-----ETEISTIQKLEAINDKLQN-----KIQEIYEKEEQLN 1535
Qy 407 -AYLEAIRKNLEWKHKDKGNKEDYD-----SKRDPINKQADAYVEKGILDKKEAEAI 461
Db 1536 IKQISEVENVNELKQFKEHRKAKDSALQSTESKMLLELTNRLQESQEEIQIMIKEK-BEM 1594
Qy 462 KRIYSSL 468
Db 1595 KRVQEAL 1601
RESULT 9
US-09-914-259-17
; Sequence 17, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-914-259-17
Query Match 6.8%; Score 163.5; DB 4; Length 956;
Best Local Similarity 20.7%; Pred. No. 0.00072;
Matches 106; Conservative 91; Mismatches 169; Indels 147; Gaps 25;
Qy 25 GSQDKSLHNR-----ELSAERPLNEQIAEEDKIKTYPPENKPGOSNYFVDNMLNLL- 78
Db 321 GORAKTIKNTVSNVLELTAE-----EWKKYKEKEKNALKSVLQHLEWEL 367
Qy 79 -----KAITEKIE-KERQSIKSSPLDNKLVED-----VDSTKNKCLDDYDSTKSL 127
Db 368 NRWNGEAVPEDEQISAKDHKSL--EPCDTPIIDNITPVVDGISAEB--EKYDEEITSL 423
Qy 128 DHKFPDPPDGLHQLDGTPLTAEDIVHKIAARIYEEND-----RAVFDKIVSKLINLGLIT 182
Db 424 YRQLDDKDDIENQ-----QSQAELKQKQMLDQDELLASTRADRYEKIQBELTRLQIEN 476
Qy 183 ESQAHTLDEVAEVLQKLSKEANNYE--EDNKPITSWTENQAGKIPEKTPMAAIQDG 239
Db 477 EA-----AKDEVEKVLQAL-BELAVNYDQKSQVEDKTRANEQLTDLAQKTTTLTTTQRE 531
Qy 240 LAK-----GENDETVSNTLTNLNGLERR-----TKT 265
Db 532 LSQQLSNHKKRATETINLLKDLGEIGIGTND-----VKTLADVNGVIEEFTWARL 588
Qy 266 YSDNFEELQYFNFVALLKS--IDSEKAKEKETLITIMKTLDIFVKMMVYGTISPEE 323
Db 589 YISKMSEVKSLVNRSKQLESQMSDNKRWASERELAAQCQLLISQHEAKISLT----- 643
Qy 324 GVSYLENLDMIALQTKNLEK-----NATDNISKLPAPSKSHEET--DSTKEBAKMEK 378
Db 644 --DYMQME-----QKRRQLESQDSLSEELAKL--RAQEKMEHVSFQDKKEKHLTRLQ- 693
Qy 379 EYGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNIEWLKHKDKGNKEDYDLSKORD 438
Db 694 -----DAEEVYKALEQQMESHR-----EAHQKQ-----LSRLRD 722

Qy 439 FINKQADAYVEKGILDK-----FAEAKRIYSSL 468
Db 723 EIEKQRIIDIRDLNQKLEQEERLSSDYNKL 755
RESULT 10
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David B.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; TITLE OF INVENTION: PORTIONS AND PRODUCTS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8991 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-714-741-32
Query Match 6.8%; Score 162; DB 4; Length 8991;
Best Local Similarity 18.2%; Pred. No. 0.02;
Matches 88; Conservative 110; Mismatches 184; Indels 102; Gaps 19;
Qy 26 SQDKSLHNRSLAERPLNEQIAEEDKIKTYPPENKPGOSNYFVDNMLNLLKAITEKE 85
Db 2024 AOEKADYQRIIEKAKEQASLEQOEANKYQLKLYLDGRNLNSSLVKKEMEAE 2083
Qy 86 KIKERQSIKSSPLDNKL-----NVEDVDSTKNKCLDDYDSTKSLDHLKFDOD----- 133
Db 2084 KDKKEQA-----EFNKIRREIVVPNPQLELMARKSEV--VKAKESGLVKRVEAEKKV 2136
Qy 134 -----DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS-KLINLGLITESQ 185
Db 2137 TEARQKLAERAKEEVLPTRVENEVHKLQKQNKK--KNILTSLASVALGAGLVTSPQ 2194
Qy 186 AHTELEVAEVLQK-----LISKEANN-----YBEDNKPPTSMT 219
Db 2195 TFVRAEESQVWEKSLEKKEBAKADTAKKDYETAKKAEADQKKYBDQKR-----T 2250

[illegible]

Qy	84	KEKLEKQIR---SSPLNKNVDEDVSTKN-----RKLDYDSTKSGLDH---KF 133
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	216	LRQLEERNLQDQDDEMEAKQNLERHISTLNQLSDSKKKLQDFASTVEALEEGKKRF 275
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	132	QDQDPGHLQDGTPLTAEIVHKTAARIYENDRAVFD-----KIVSKLLNLGLITESQA 186
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	276	QKEIENUTQYEEKAAAYDKLEKTKNRLQQLDLDLVLDNQRQLVSNL-----EKKQ 328
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	187	HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTTPMA-AIQDGLAKGEN 245
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	329	RKPDQLAAE-EKNIS---SKYADERDR---AEAAREKETKALSLARALEEALEAKEE 379
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	246	DETVSNLTLTNGLERTKTYSED-----NPELOYPFNFYALLKSIDSEKAKEKETLIT 301
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	380	LERTNMLK---AEMEDLVSSKDDVGKNVHEL-----EKSRALETQME 420
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	302	IMKTLIDFVMMVKYGTISPPEGVSYLENLDEMIALQTKNKLKNATDNISK-----L 354
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	421	EMKTLQLESEDELQ-----ASEDAKLRLVNMQAUKGOFERDLQARDEQNEKRRLOLQRL 476
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	355	FPAPSEKSHETSTKBEAAKMEKEYGSLKSDTDDNSNPGCKTDEPK--GKTEAYLEAI 412
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	477	HEYETELEDERNERALAAAKKLE-GDLKDLQADSAIKGREEAIKQLRKLQAKMKDF 535
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	413	RKNIE-----WLKXHKGNKEDYDLSKWRDPI-----NKQADAYVEKGILDK 455
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db	536	QRELEDARASRDEIFATAKENEKKAKSLSLADLMQLQEDLAAARARQAD-----LEK 588
		:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy	456	EE 457
Db	589	EE 590
RESULT 12		
US-08-742-923A-6		
; Sequence 6, Application US/08742923A		
; Patent No. 5869611		
; GENERAL INFORMATION:		
; APPLICANT: Liu, Pu		
; APPLICANT: Collins, Francis S.		
; APPLICANT: Siciliano, Michael J.		
; APPLICANT: Claxton, David		
; TITLE OF INVENTION: Markers for Detection of Chromosome 16		
; TITLE OF INVENTION: Rearrangements		
; NUMBER OF SEQUENCES: 14		
; CORRESPONDENCE ADDRESS:		
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.		
; STREET: P.O. Box 828		
; CITY: Bloomfield Hills		
; STATE: MI		
; COUNTRY: USA		
; ZIP: 48303		
; COMPUTER READABLE FORM:		
; MEDIUM TYPE: Floppy disk		
; COMPUTER: IBM PC compatible		
; OPERATING SYSTEM: PC-DOS/MS-DOS		
; SOFTWARE: PatentIn Release #1.0, Version #1.25		
; CURRENT APPLICATION DATA:		
; APPLICATION NUMBER: US/08/742,923A		
; FILING DATE: No. 5869611member 1, 1996		
; CLASSIFICATION: 435		
; ATTORNEY/AGENT INFORMATION:		
; NAME: Smith, Deann F.		
; REGISTRATION NUMBER: 36683		
; REFERENCE/DOCKET NUMBER: 2115-00869DVC		
; TELECOMMUNICATION INFORMATION:		
; TELEPHONE: (810) 641-1600		
; TELEFAX: (810) 641-0270		
; INFORMATION FOR SEQ ID NO: 6:		
; SEQUENCE CHARACTERISTICS:		
; LENGTH: 816 amino acids		
; TYPE: amino acid		
; TOPOLOGY: linear		

MOLECULE TYPE: protein
US-08-742-923A-6

Query Match 6.6%; Score 159.5; DB 2; Length 816;
Best Local Similarity 22.6%; Pred. No. 0.0011; Indels 99; Gaps 21;

Matches 109; Conservative 81; Mismatches 193; Mismatches 190; Indels 99; Gaps 21;

QY 26 SQDKSLHRLSARPLNEQIAEAEEDKIKKTPPENKPGQSNYSFV--DNLNLLKAITKE 83
DB 158 SHREMEVESVTCMLNE--AEGKAIKAKDVASLSQLODTQELLQEEETROKLNVTK 215
QY 84 KEKTEKQSTR---SSPLDNKLNVEDVDSTKN-----RKLIDDYDSTKGLDH---KF 131
DB 216 LRQLEERNLSQDQDEEMAKQNLERHISTLNQLSDSKKKLQDFASTVEALEEGKRF 275
QY 132 QDDPDGLHQLDGTPLTAEDIVHKAARIYEENDRAVFD-----KIVSKLNLGLITESQA 186
DB 276 QKEIENLTQQVEEKAAYDKLEKTKNRLOQELDLVVDLNDQRLVSNL-----EKKQ 328
QY 187 HTLEDEVAELQKLISKEANNVEEDPNKPTSWTENQAGKIPEKVTPTMA-AIODGLAKGEN 245
DB 329 RKPQLLAE--EKNI---SKYADERDR---AEAEAREKETKALSARALEEAEKEE 379
QY 246 DETVSNLTLTNGLERRTKYSED-----NPEELQYFPNFPYALLKSIDSEKEAKEKETLIT 301
DB 380 LERTNKMK---AEMEDLVSSKDDVGNVHEL-----EKSRALETQME 420
QY 302 IMKTLIDFVKMVKYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISK-----L 354
DB 421 EMKTLQEELELQ-----ASEDAKLRLEVNMQALKQGFERDLQARDEQNEEKRLQRL 476
QY 355 FPAPSEKSHETDSTKEAAKMEKEYGSLKSDTKDSDNSPGKTDDEPK--GKTEAYLEAI 412
DB 477 HEYETELEDERNERALAAAKKLE-GDLKLELQADSIAKGREAIKQLRKLQAKMDF 535
QY 413 RKNIE-----WLKHKDKGNKEDYDLSKMRDPI-----NKQADAYVEKGILDK 455
DB 536 QRELEDAARSDEIFATAKENEKALEADLMQLQEDLAAAEARAKQAD-----LEK 588
QY 456 EE 457
DB 589 EE 590

RESULT 13

US-08-533-306A-4
Sequence 4, Application US/08533306A
Patent No. 5837457

GENERAL INFORMATION:
APPLICANT: Liu, Pu

APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.

APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16

TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
ADDRESSEE: Harnes, Dickey & Pierce, P.L.C.

STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI

COUNTRY: USA
ZIP: 48303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/533.306A
FILING DATE: September 25, 1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:

TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 885 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-533-306A-4

Query Match 6.6%; Score 158.5; DB 2; Length 885;
Best Local Similarity 22.9%; Pred. No. 0.0015; Indels 99; Gaps 21;

Matches 109; Conservative 79; Mismatches 190; Indels 99; Gaps 21;

QY 31 LHNRELSAERPLNEQIAEAEEDKIKKTPPENKPGQSNYSFV--DNLNLLKAITKEKIE 88
DB 232 LONEVESVTCMLNE--AEGKAIKAKDVASLSQLODTQELLQEEETROKLNVTKRLQLE 289
QY 89 KEROSIR---SSPLDNKLNVEDVDSTKN-----RKLIDDYDSTKGLDH---KFQDDPD 136
DB 290 EERNLSQDQDEEMAKQNLERHISTLNQLSDSKKKLQDFASTVEALEEGKRFQKEIE 349
QY 137 GLHOLDGTPLTAEDIVHKAARIYEENDRAVFD-----KIVSKLNLGLITESQAHTLED 191
DB 350 NLTOQEKAAYDKLEKTKNRLOQELDLVVDLNDQRLVSNL-----EKKQKFPQ 402
QY 192 EVASVLQKLISKEANNVEEDPNKPTSWTENQAGKIPEKVTPTMA-AIODGLAKGENDETVS 250
DB 403 LLAE--EKNI---SKYADERDR---AEAEAREKETKALSARALEEAEKEELERN 453
QY 251 NTLTLTNGLERRTKYSED-----NPEELQYFPNFPYALLKSIDSEKEAKEKETLITIMTKL 306
DB 454 KMLK---AEMEDLVSSKDDVGNVHEL-----EKSRALETQMEEMKTQ 494
QY 307 IDFKVMVKYGTISPPEGVSYLENLDEMIALQTKNLEKNATDNISK-----LPPAPS 359
DB 495 LEELEDELQ-----ASEDAKLRLEVNMQALKQGFERDLQARDEQNEEKRLQRLQHEVET 550
QY 360 EKSHEETDSTKEAAKMEKEYGSLKSDTKDSDNSPGKTDDEPK--GKTEAYLEAIRKNIE 417
DB 551 ELEDERNERALAAAKKLE-GDLKLELQADSIAKGREAIKQLRKLQAKMDFQRELE 609
QY 418 -----WLKHKDKGNKEDYDLSKMRDPI-----NKQADAYVEKGILDKIE 457
DB 610 DARASDEIFATAKENEKALEADLMQLQEDLAAAEARAKQAD-----LEKEE 659

RESULT 14

US-08-742-923A-4
Sequence 4, Application US/08742923A
Patent No. 5869611

GENERAL INFORMATION:
APPLICANT: Liu, Pu

APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.

APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16

TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
ADDRESSEE: Harnes, Dickey & Pierce, P.L.C.

STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI

COUNTRY: USA
ZIP: 48303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742.923A
; FILING DATE: No. 5869611member 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-742-923A-4

Query Match      6.6%; Score 158.5; DB 2; Length 885;
Best Local Similarity 22.9%; Pred. No. 0.0015;
Matches 109; Conservative 79; Mismatches 190; Indels 99; Gaps 21;

Qy 31 LHNRELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSFV--DNLNLLKAITKEKIE 88
Db 232 LQNEVESVTGMLNE--AEGBAIIKAKDVASLSQLODQELLQEBETROKLVNSTKLRQLE 289

Qy 89 KEROSIR---SSPLDNKLVNEDVSTKN-----RKLDDYDSTKSGLDH---KFQDDPD 136
Db 290 EERNSLQQLDEEMAKONLHERHSTLNIQLSDSKKKLQDFASTVEALEGKKRFQKEIE 349

Qy 137 GLHQLDGTPLTAEDIVHKIAAIIYEENDRAVFD-----KIVSKLNLNGLITESQAHTLED 191
Db 350 NLTQYEEKAAAYDKLETKNRLQQLDLDLVVDLDNQRLVSNL-----EKKQKEDQ 402

Qy 192 EVAEVLQKLIKEANNVEEDPNKPTSWTENQAGKIPEKVTMA-AIQDGLAKGENDETVS 250
Db 403 LLAE--EKNIS---SKYADERDR---AAEAREKETKALSARALBEALEAKEELERTN 453

Qy 251 NTLTLTNGLERRTYSED---NFEELQYFPNFYALLKSIDSEKAKEKTLITIMKTL 306
Db 454 KQLK-----AEMEDLVSSKDDVGKVNHEL-----BKSRALETQMEEMKTQ 494

Qy 307 IDFKVMVKYGTISPFGVSYLENLDEMIALQTKNLEKNATDNISK-----LFPAPS 359
Db 495 LEELEDELQ---ASEDAKLELVNMQALKQCFERDQARDEQNEEKRQRLQRLQHEYT 550

Qy 360 EKSHEETDSTKEAAKMEKEYGSLKSDTKDSDNSPGGKTIDEPK--GKTEAYLEAIRKNIE 417
Db 551 ELEDERNERALAAAKKLE--GDLKDLQADSIAIKGREBAIKQLRKLQAKMDFQRELE 609

Qy 418 -----WLKHKDKGKNKEDYDLSKMRDFI-----NKQADAVVEKGILDKEE 457
Db 610 DARASRDEIFATAKENEKAKSLKLEADLMQLQEDLAAAARAKQAD-----LEKEE 659
```

RESULT 15

```
US-09-914-259-12
; Sequence 12, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
```

```
; LENGTH: 1087
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; US-09-914-259-12

Query Match      6.5%; Score 156.5; DB 4; Length 1087;
Best Local Similarity 19.3%; Pred. No. 0.0028;
Matches 103; Conservative 83; Mismatches 180; Indels 167; Gaps 22;

Qy 31 LHNRELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSFVDNLNLLKAITKEKIEKE 90
Db 304 IHEKEVEIDR-LNEQIIKLOQ-----LKITTDKNVIEEK 337

Qy 91 QRSIRSSPLDNKLVNEDVST-KNR-----KLDDYDSTKSGLDHKKFDDPDGLHQL 141
Db 338 NELIRDLEAQIECLMSQDERVRKNREBEIEQLNEVIEKIQELANIQKTSVDPSSLSE- 396

Qy 142 DGTPLTAEDIVHKIAAIIYEENDRAVFDKIVSKLNLNGLITESQAHTLEDVA----- 194
Db 397 -----EADSLKHQL-----DKVIAEKAL-----EHQVETTNNEEMAVTKNVLK 434

Qy 195 -----EVLQKLIS--KEANNVEEDPNKPTSWTENQAGKIPEKVTMAAI--ODGLAKG 243
Db 435 ETNPKMQLTQELCSLKREREKMERIQSVPEKSVNMSVGLSKDKKPEMDLIPTEDALAQL 494

Qy 244 E-----NDETVSNTL-TLTNGLERRTKYSEDNFEELQYFPNFYALLKSIDSEKAKE 295
Db 495 ETOTQLRSSESSKVSLSLETKLQLESTVSTKDLTLTCY-----KIQDMREQGR 547

Qy 296 KET-----LITIMKTL-----IDFKVMVKYGTISPFG- 324
Db 548 SETEMLTQKIVSLQKVLVEEKVAAALVSQVQLAEVQYVVKLCADKPAVSSDPARTEVPGLS 607

Qy 325 -----VSYLENL-----DEMIALQTKNLEK-NATDNISKLF 355
Db 608 QLAGNTVESDVSAITWRISELSQVEMHSLISEKEQVEIAEKNALEKEKQLQELQKLV 667

Qy 356 PAPSEKSHETDSTKEAAKMEKEYGSLKSDTKDSDNSNPG-----GKTDEPKGKTAYL 409
Db 668 QDSETKQRE-----RERQSRHLGDLGVLESTTSSESGVFGLEALRAESAAPKGLANYK 722

Qy 410 EATRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADA---YVEKGILDKEEAE 459
Db 723 ELAEKLOEELLVKETNNASLPKLSLHVDRDQLEAEDKLSHFSEKE--DKTEVQ 773
```

Search completed: January 2, 2004, 11:03:01

Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 11:00:07 ; Search time 38 Seconds
(without alignments)
2471.055 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGWLVLVLPQAF.....EKGILDKKEAEAKRIYSSL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2399	100.0	468	11	US-09-931-836-67
2	2399	100.0	468	12	US-10-035-977-67
3	2399	100.0	468	12	US-10-063-735-150
4	2399	100.0	468	12	US-10-137-870-90
5	2399	100.0	468	12	US-10-140-018-90
6	2399	100.0	468	12	US-10-140-021-90
7	2399	100.0	468	12	US-10-140-274-90
8	2399	100.0	468	12	US-10-140-471-90
9	2399	100.0	468	12	US-10-140-807-90
10	2399	100.0	468	12	US-10-140-922-90
11	2399	100.0	468	12	US-10-140-924-90
12	2399	100.0	468	12	US-10-140-926-90
13	2399	100.0	468	12	US-10-141-698-90
14	2399	100.0	468	12	US-10-141-702-90
15	2399	100.0	468	12	US-10-141-704-90

16	2399	100.0	468	12	US-10-142-421-90	Sequence 90, Appl
17	2399	100.0	468	12	US-10-142-432-90	Sequence 90, Appl
18	2399	100.0	468	12	US-10-142-767-90	Sequence 90, Appl
19	2399	100.0	468	12	US-10-143-033-90	Sequence 90, Appl
20	2399	100.0	468	12	US-10-144-994-90	Sequence 90, Appl
21	2399	100.0	468	12	US-10-145-628-90	Sequence 90, Appl
22	2399	100.0	468	12	US-10-145-631-90	Sequence 90, Appl
23	2399	100.0	468	12	US-10-145-633-90	Sequence 90, Appl
24	2399	100.0	468	12	US-10-145-746-90	Sequence 90, Appl
25	2399	100.0	468	12	US-10-145-748-90	Sequence 90, Appl
26	2399	100.0	468	12	US-10-145-823-90	Sequence 90, Appl
27	2399	100.0	468	12	US-10-145-826-90	Sequence 90, Appl
28	2399	100.0	468	12	US-10-145-870-90	Sequence 90, Appl
29	2399	100.0	468	12	US-10-145-876-90	Sequence 90, Appl
30	2399	100.0	468	12	US-10-145-959-90	Sequence 90, Appl
31	2399	100.0	468	12	US-10-146-724-90	Sequence 90, Appl
32	2399	100.0	468	12	US-10-146-725-90	Sequence 90, Appl
33	2399	100.0	468	12	US-10-146-795-90	Sequence 90, Appl
34	2399	100.0	468	12	US-10-147-495-90	Sequence 90, Appl
35	2399	100.0	468	12	US-10-147-501-90	Sequence 90, Appl
36	2399	100.0	468	12	US-10-147-504-90	Sequence 90, Appl
37	2399	100.0	468	12	US-10-147-506-90	Sequence 90, Appl
38	2399	100.0	468	12	US-10-147-509-90	Sequence 90, Appl
39	2399	100.0	468	12	US-10-147-510-90	Sequence 90, Appl
40	2399	100.0	468	12	US-10-147-511-90	Sequence 90, Appl
41	2399	100.0	468	12	US-10-147-529-90	Sequence 90, Appl
42	2399	100.0	468	12	US-10-152-397-90	Sequence 90, Appl
43	2399	100.0	468	12	US-10-153-586-90	Sequence 90, Appl
44	2399	100.0	468	12	US-10-158-783-90	Sequence 90, Appl
45	2399	100.0	468	12	US-10-158-786-90	Sequence 90, Appl

ALIGNMENTS

RESULT 1

US-09-931-836-67
; Sequence 67, Application US/09931836
; Publication No. US20030027249A1
; GENERAL INFORMATION:
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3030R1C1
; CURRENT APPLICATION NUMBER: US/09/931,836
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/112514
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114140
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115552
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22

;; PRIOR APPLICATION NUMBER: 60/125774
;; PRIOR FILING DATE: 1999-03-23
;; PRIOR APPLICATION NUMBER: 60/125778
;; PRIOR FILING DATE: 1999-03-23
;; PRIOR APPLICATION NUMBER: 60/125826
;; PRIOR FILING DATE: 1999-03-24
;; PRIOR APPLICATION NUMBER: 60/127035
;; PRIOR FILING DATE: 1999-03-31
;; PRIOR APPLICATION NUMBER: 60/127706
;; PRIOR FILING DATE: 1999-04-05
;; PRIOR APPLICATION NUMBER: 60/129122
;; PRIOR FILING DATE: 1999-04-13
;; PRIOR APPLICATION NUMBER: 60/130359
;; PRIOR FILING DATE: 1999-04-21
;; PRIOR APPLICATION NUMBER: 60/131270
;; PRIOR FILING DATE: 1999-04-27
;; PRIOR APPLICATION NUMBER: 60/131272
;; PRIOR FILING DATE: 1999-04-27
;; PRIOR APPLICATION NUMBER: 60/131291
;; PRIOR FILING DATE: 1999-04-27
;; PRIOR APPLICATION NUMBER: 60/132371
;; PRIOR FILING DATE: 1999-05-04
;; PRIOR APPLICATION NUMBER: 60/132379
;; PRIOR FILING DATE: 1999-05-04
;; PRIOR APPLICATION NUMBER: 60/132383
;; PRIOR FILING DATE: 1999-05-04
;; PRIOR APPLICATION NUMBER: 60/135750
;; PRIOR FILING DATE: 1999-05-25
;; PRIOR APPLICATION NUMBER: 60/138166
;; PRIOR FILING DATE: 1999-06-08
;; PRIOR APPLICATION NUMBER: 60/144791
;; PRIOR FILING DATE: 1999-07-20
;; PRIOR APPLICATION NUMBER: 60/146970
;; PRIOR FILING DATE: 1999-08-03
;; PRIOR APPLICATION NUMBER: 60/162506
;; PRIOR FILING DATE: 1999-10-29
;; PRIOR APPLICATION NUMBER: 09/311832
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 09/380142
;; PRIOR FILING DATE: 1999-08-25
;; PRIOR APPLICATION NUMBER: 09/644848
;; PRIOR FILING DATE: 2000-08-22
;; PRIOR APPLICATION NUMBER: 09/747259
;; PRIOR FILING DATE: 2000-12-20
;; PRIOR APPLICATION NUMBER: 09/816744
;; PRIOR FILING DATE: 2001-03-22
;; PRIOR APPLICATION NUMBER: 09/854208
;; PRIOR FILING DATE: 2001-05-10
;; PRIOR APPLICATION NUMBER: 09/854280
;; PRIOR FILING DATE: 2001-05-10
;; PRIOR APPLICATION NUMBER: 09/874503
;; PRIOR FILING DATE: 2001-06-05
;; PRIOR APPLICATION NUMBER: 09/869599
;; PRIOR FILING DATE: 2001-06-29
;; PRIOR APPLICATION NUMBER: 09/908,827
;; PRIOR FILING DATE: 2001-07-18
;; PRIOR APPLICATION NUMBER: PCT/US99/10733
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: PCT/US99/28551
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30720
;; PRIOR FILING DATE: 1999-12-22
;; PRIOR APPLICATION NUMBER: PCT/US00/05601
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: PCT/US00/05841
;; PRIOR FILING DATE: 2000-03-02
;; PRIOR APPLICATION NUMBER: PCT/US00/14042
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: PCT/US00/15264
;; PRIOR FILING DATE: 2000-06-02
;; PRIOR APPLICATION NUMBER: PCT/US00/23522
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: PCT/US00/23328

;; PRIOR FILING DATE: 2000-08-24
;; PRIOR APPLICATION NUMBER: PCT/US00/32678
;; PRIOR FILING DATE: 2000-12-01
;; PRIOR APPLICATION NUMBER: PCT/US00/34956
;; PRIOR FILING DATE: 2000-12-20
;; PRIOR APPLICATION NUMBER: PCT/US01/06520
;; PRIOR FILING DATE: 2001-08-28
;; PRIOR APPLICATION NUMBER: PCT/US01/17800
;; PRIOR FILING DATE: 2001-06-01
;; PRIOR APPLICATION NUMBER: PCT/US01/19692
;; PRIOR FILING DATE: 2001-06-20
;; PRIOR APPLICATION NUMBER: PCT/US01/21066
;; PRIOR FILING DATE: 2001-06-29
;; PRIOR APPLICATION NUMBER: PCT/US01/21735
;; PRIOR FILING DATE: 2001-07-09
;; NUMBER OF SEQ ID NOS: 80
;; SEQ ID NO 67
;; LENGTH: 468
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-931-836-67

Query Match 100.0%; Score 2399; DB 11; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGFLGTGWLVLVLPICAFPKPGSQSKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
DB 1 MGFLGTGWLVLVLPICAFPKPGSQSKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

QY 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIEROSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
DB 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIEROSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

QY 121 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDI VHKAARIYEENDRAVFDKIVSKLLNLGL 180
DB 121 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDI VHKAARIYEENDRAVFDKIVSKLLNLGL 180

QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKFTSWTENQAGKIPKVTMAAIQDGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKFTSWTENQAGKIPKVTMAAIQDGL 240

QY 241 AKGENDETVSNTLTITNGLERRTKTYSSEDFEELQYFPNFVALLKSIDSEKEAKEKETLI 300
DB 241 AKGENDETVSNTLTITNGLERRTKTYSSEDFEELQYFPNFVALLKSIDSEKEAKEKETLI 300

QY 301 TIMKTLIDFVQMMVKYGTISP EEGVSYLENDEMIALQTKNKLKXNATDNTSKLPAPSE 360
DB 301 TIMKTLIDFVQMMVKYGTISP EEGVSYLENDEMIALQTKNKLKXNATDNTSKLPAPSE 360

QY 361 KSHEETDSTKEEAAKWEKEYGSLKDDSTKDDNSPGKTDPEPKGTEAYLEAIRKNIEWLK 420
DB 361 KSHEETDSTKEEAAKWEKEYGSLKDDSTKDDNSPGKTDPEPKGTEAYLEAIRKNIEWLK 420

QY 421 KHKKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
DB 421 KHKKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468

RESULT 2
US-10-035-977-67
;; Sequence 67, Application US/10035977
;; Publication No. US20030134327A1
;; GENERAL INFORMATION:
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Eaton, Dan L.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.

APPLICANT: Zhang Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3030R1C10
CURRENT APPLICATION NUMBER: US/10/035,977
CURRENT FILING DATE: 2001-12-26
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/112514
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113300
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113430
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114140
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115552
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/125774
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125826
PRIOR FILING DATE: 1999-03-24
PRIOR APPLICATION NUMBER: 60/127035
PRIOR FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/127706
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/129122
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 60/130359
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131272
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/132371
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132379
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132383
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138166
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/144791
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/146970
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380142
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/644848
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 09/747259
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/816744
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/854208
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/854280

PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/874503
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: 09/869599
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/908,827
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US99/10733
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30720
PRIOR FILING DATE: 1999-12-22
PRIOR APPLICATION NUMBER: PCT/US00/05601
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: PCT/US00/34956
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/17800
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: PCT/US01/19692
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: PCT/US01/21066
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: PCT/US01/21735
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 80
SEQ ID NO 67
LENGTH: 468
TYPE: PRT
ORGANISM: Homo Sapien
US-10-035-977-67

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8,7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTVILVLVLPQAPFKPGGQDKSLHNRLSAERPLNEQIAEAEEDKIKKTYPP 60
Db 1 MGFLGTGTVILVLVLPQAPFKPGGQDKSLHNRLSAERPLNEQIAEAEEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKAI TEKEKIERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy 121 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHKKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180

Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKPEKVTTPMAAIODGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEDPNKPTSWTENQAGKPEKVTTPMAAIODGL 240

Qy 241 AKGENDETQSNLTTLTNGLERRTKTYSEDNFEELOQFPNFPYALLKSIDSEKAEKETLI 300
Db 241 AKGENDETQSNLTTLTNGLERRTKTYSEDNFEELOQFPNFPYALLKSIDSEKAEKETLI 300

Qy 301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNLEKNATDNISKLFAPAPSE 360
Db 301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNLEKNATDNISKLFAPAPSE 360

; Sequence 90, Application US/10140274
; Publication No. US20030143674A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C161
; CURRENT APPLICATION NUMBER: US/10/140,274
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-274-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGTLVILVLPPIQAPFKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
DB 1 MGFLGTGTLVILVLPPIQAPFKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
QY 61 ENKPGQSNYSFVDNLLKKAITEKEIEKERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
DB 61 ENKPGQSNYSFVDNLLKKAITEKEIEKERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
QY 121 DSTKSGLDHFKQDDPGDLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVSKLLNLGL 180
DB 121 DSTKSGLDHFKQDDPGDLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVSKLLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIODGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIODGL 240
QY 241 AKGNDSTVNTLTNLGLRRTKTYSEDNPEELQYFPNFPYALLKSIDSEKEAKEKETLI 300
DB 241 AKGNDSTVNTLTNLGLRRTKTYSEDNPEELQYFPNFPYALLKSIDSEKEAKEKETLI 300
QY 301 TIMKTLIDFVKMVKYCTISPEEGSVYLENDEMIALQTKNLEKNATDNISKLFPAPSE 360
DB 301 TIMKTLIDFVKMVKYCTISPEEGSVYLENDEMIALQTKNLEKNATDNISKLFPAPSE 360
QY 361 KSHEETDSTKEAAKMEKEYGSLKSDTSDNSPGGKTDEPKGTAYLEAIRKNIEWLK 420
DB 361 KSHEETDSTKEAAKMEKEYGSLKSDTSDNSPGGKTDEPKGTAYLEAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGLDKKEAEAIKRIYSSL 468
DB 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGLDKKEAEAIKRIYSSL 468

RESULT 8
US-10-140-471-90
; Sequence 90, Application US/10140471

; Publication No. US2003013887A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C163
; CURRENT APPLICATION NUMBER: US/10/140,471
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGFLGTGTLVILVLPPIQAPFKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
DB 1 MGFLGTGTLVILVLPPIQAPFKPGSQDQKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
QY 61 ENKPGQSNYSFVDNLLKKAITEKEIEKERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
DB 61 ENKPGQSNYSFVDNLLKKAITEKEIEKERQSRSSPLDNKLNVEDVDSTKNRKLIDDY 120
QY 121 DSTKSGLDHFKQDDPGDLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVSKLLNLGL 180
DB 121 DSTKSGLDHFKQDDPGDLHQLDGTPLTAEDIVHKAARIYEENDRAVFDKIVSKLLNLGL 180
QY 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIODGL 240
DB 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIODGL 240
QY 241 AKGNDSTVNTLTNLGLRRTKTYSEDNPEELQYFPNFPYALLKSIDSEKEAKEKETLI 300
DB 241 AKGNDSTVNTLTNLGLRRTKTYSEDNPEELQYFPNFPYALLKSIDSEKEAKEKETLI 300
QY 301 TIMKTLIDFVKMVKYCTISPEEGSVYLENDEMIALQTKNLEKNATDNISKLFPAPSE 360
DB 301 TIMKTLIDFVKMVKYCTISPEEGSVYLENDEMIALQTKNLEKNATDNISKLFPAPSE 360
QY 361 KSHEETDSTKEAAKMEKEYGSLKSDTSDNSPGGKTDEPKGTAYLEAIRKNIEWLK 420
DB 361 KSHEETDSTKEAAKMEKEYGSLKSDTSDNSPGGKTDEPKGTAYLEAIRKNIEWLK 420
QY 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGLDKKEAEAIKRIYSSL 468
DB 421 KHDKGNKEDYDLSKMRDFINKQADAYVEKGLDKKEAEAIKRIYSSL 468

RESULT 9
US-10-140-807-90
; Sequence 90, Application US/10140807
; Publication No. US20030134354A1

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C174
; CURRENT APPLICATION NUMBER: US/10/140.807
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-807-90

```

```

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTWILVLPIQAPFKPGSQDSKLNHRELSAERPLNEQIAEAEEDKIKTYPP 60
Db 1 MGFLGTGTWILVLPIQAPFKPGSQDSKLNHRELSAERPLNEQIAEAEEDKIKTYPP 60
Qy 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
Qy 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIODGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIODGL 240
Qy 241 AKGENDETVSNTLTITNGLERRTKTYSNDNFEELOYPNFPYALLKSIDSEKEAKEKETLI 300
Db 241 AKGENDETVSNTLTITNGLERRTKTYSNDNFEELOYPNFPYALLKSIDSEKEAKEKETLI 300
Qy 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNKLKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNKLKNATDNISKLPAPSE 360
Qy 361 KSHETDSTKEAAKMEKEYGSLKSDTKDNNPCKGTDPEPKGTEAYLEAIRKNIEWLK 420
Db 361 KSHETDSTKEAAKMEKEYGSLKSDTKDNNPCKGTDPEPKGTEAYLEAIRKNIEWLK 420
Qy 421 KHKKGKNGKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAKRIYSSL 468
Db 421 KHKKGKNGKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAKRIYSSL 468

```

```

RESULT 10
US-10-140-922-90
; Sequence 90, Application US/10140922
; Publication No. US2003013889A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C179
; CURRENT APPLICATION NUMBER: US/10/140.922
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-922-90

```

```

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTWILVLPIQAPFKPGSQDSKLNHRELSAERPLNEQIAEAEEDKIKTYPP 60
Db 1 MGFLGTGTWILVLPIQAPFKPGSQDSKLNHRELSAERPLNEQIAEAEEDKIKTYPP 60
Qy 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKAITKEKIEKERQSRSSPLDNKLNVEDVDSTKRNKLIDDY 120
Qy 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Qy 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIODGL 240
Db 181 ITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVTTPMAAIODGL 240
Qy 241 AKGENDETVSNTLTITNGLERRTKTYSNDNFEELOYPNFPYALLKSIDSEKEAKEKETLI 300
Db 241 AKGENDETVSNTLTITNGLERRTKTYSNDNFEELOYPNFPYALLKSIDSEKEAKEKETLI 300
Qy 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNKLKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNKLKNATDNISKLPAPSE 360
Qy 361 KSHETDSTKEAAKMEKEYGSLKSDTKDNNPCKGTDPEPKGTEAYLEAIRKNIEWLK 420
Db 361 KSHETDSTKEAAKMEKEYGSLKSDTKDNNPCKGTDPEPKGTEAYLEAIRKNIEWLK 420
Qy 421 KHKKGKNGKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAKRIYSSL 468
Db 421 KHKKGKNGKEDYDLSKMRDFINKQADAYVEKGILDKKEAEAKRIYSSL 468

```

```

RESULT 11
US-10-140-924-90
; Sequence 90, Application US/10140924
; Publication No. US20030134355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```

```

; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C177
; CURRENT APPLICATION NUMBER: US/10/140,924
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-924-90

```

```

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8,7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTLVLVLPQAPKPGSQDSKLSHNSLSAERPLNEQIAEAEEDKIKKTYPP 60
Db 1 MGFLGTGTLVLVLPQAPKPGSQDSKLSHNSLSAERPLNEQIAEAEEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLLKAITTEKEKIEKERSQSRSSPLDNKLVNVEDVSTKRNKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKAITTEKEKIEKERSQSRSSPLDNKLVNVEDVSTKRNKLIDDY 120

Qy 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180

Qy 181 ITESQAHLEDEVAEVLQKLISKEANNYEDPNKPTSWTENQAGKIPKVTMPMAIIOGGL 240
Db 181 ITESQAHLEDEVAEVLQKLISKEANNYEDPNKPTSWTENQAGKIPKVTMPMAIIOGGL 240

Qy 241 AKGNDSTVNTLTTLTNGLERRTKTSYSEDNFEELQYFPNFYALLKSIDSEKAEKETLI 300
Db 241 AKGNDSTVNTLTTLTNGLERRTKTSYSEDNFEELQYFPNFYALLKSIDSEKAEKETLI 300

Qy 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360

Qy 361 KSHEETDSTKEEAAKEKEYGSLKSDTKDONSPPGKTDSPKGTAEYLAIRKNIEWLK 420
Db 361 KSHEETDSTKEEAAKEKEYGSLKSDTKDONSPPGKTDSPKGTAEYLAIRKNIEWLK 420

Qy 421 KHKDKGNKEDYDLSKMRDFINKQADAVVEKGLDKEEAEAIKRIYSSL 468
Db 421 KHKDKGNKEDYDLSKMRDFINKQADAVVEKGLDKEEAEAIKRIYSSL 468

```

```

RESULT 12
US-10-140-926-90
; Sequence 90, Application US/10140926
; Publication No. US20030134356A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen

```

```

; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C187
; CURRENT APPLICATION NUMBER: US/10/140,926
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-926-90

```

```

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8,7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTLVLVLPQAPKPGSQDSKLSHNSLSAERPLNEQIAEAEEDKIKKTYPP 60
Db 1 MGFLGTGTLVLVLPQAPKPGSQDSKLSHNSLSAERPLNEQIAEAEEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLLKAITTEKEKIEKERSQSRSSPLDNKLVNVEDVSTKRNKLIDDY 120
Db 61 ENKPGQSNYSFVDNLNLLKAITTEKEKIEKERSQSRSSPLDNKLVNVEDVSTKRNKLIDDY 120

Qy 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180
Db 121 DSTKSGLDHFKQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180

Qy 181 ITESQAHLEDEVAEVLQKLISKEANNYEDPNKPTSWTENQAGKIPKVTMPMAIIOGGL 240
Db 181 ITESQAHLEDEVAEVLQKLISKEANNYEDPNKPTSWTENQAGKIPKVTMPMAIIOGGL 240

Qy 241 AKGNDSTVNTLTTLTNGLERRTKTSYSEDNFEELQYFPNFYALLKSIDSEKAEKETLI 300
Db 241 AKGNDSTVNTLTTLTNGLERRTKTSYSEDNFEELQYFPNFYALLKSIDSEKAEKETLI 300

Qy 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360
Db 301 TIMKTLIDFVKMVKYGTISPESGVSYLENDEMIALQTKNLEKNATDNISKLPAPSE 360

Qy 361 KSHEETDSTKEEAAKEKEYGSLKSDTKDONSPPGKTDSPKGTAEYLAIRKNIEWLK 420
Db 361 KSHEETDSTKEEAAKEKEYGSLKSDTKDONSPPGKTDSPKGTAEYLAIRKNIEWLK 420

Qy 421 KHKDKGNKEDYDLSKMRDFINKQADAVVEKGLDKEEAEAIKRIYSSL 468
Db 421 KHKDKGNKEDYDLSKMRDFINKQADAVVEKGLDKEEAEAIKRIYSSL 468

```

```

RESULT 13
US-10-141-698-90
; Sequence 90, Application US/10141698
; Publication No. US20030134357A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura

```


APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tamas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3330R1C209

CURRENT APPLICATION NUMBER: US/10/141,704

CURRENT FILING DATE: 2002-05-08

Prior Application removed - See Palm or File Wrapper

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 90

LENGTH: 468

TYPE: PRT

ORGANISM: Homo Sapien

US-10-141-704-90

Query Match

Best Local Similarity 100.0%; Score 2399; DB 12; Length 468;

Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGFLGTGTLVLVLPQAFPKPGSQDKSLHNRELQAERPLNEQIAEAEDKIKKTYPP 60

Db 1 MGFLGTGTLVLVLPQAFPKPGSQDKSLHNRELQAERPLNEQIAEAEDKIKKTYPP 60

Qy 61 ENKPGQSNYSFVDNLNLLKATEKEKEKQSRSSPLDNKLNVEDVDSKRNKLIDDY 120

Db 61 ENKPGQSNYSFVDNLNLLKATEKEKEKQSRSSPLDNKLNVEDVDSKRNKLIDDY 120

Qy 121 DSTKSGLDHFKQDDPDGHLQHDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180

Db 121 DSTKSGLDHFKQDDPDGHLQHDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLLNLGL 180

Qy 181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIQDGL 240

Db 181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIQDGL 240

Qy 241 AKGENDETIVNTLTNGLERRTKYSEDNFEELQYFPNFYALUKSIDSEKEAKEKETLI 300

Db 241 AKGENDETIVNTLTNGLERRTKYSEDNFEELQYFPNFYALUKSIDSEKEAKEKETLI 300

Qy 301 TIMKTLIDFVKMVKYGTISPESGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360

Db 301 TIMKTLIDFVKMVKYGTISPESGVSYLENLDEMIALQTKNLEKNATDNISKLPAPSE 360

Qy 361 KSHETDSTKEEAAKMEKEYGSLKSDKDNNSPGGKTDPEPKGTETAYLEAIRKNIEWLK 420

Db 361 KSHETDSTKEEAAKMEKEYGSLKSDKDNNSPGGKTDPEPKGTETAYLEAIRKNIEWLK 420

Qy 421 KHDKGNKEDYDLSKMRDFINKQADAVEKGILDKEEAEAIKRIYSSL 468

Db 421 KHDKGNKEDYDLSKMRDFINKQADAVEKGILDKEEAEAIKRIYSSL 468

Search completed: January 2, 2004, 11:03:51

Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 21 Seconds
(without alignments)
2143.188 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTLVLVLPQAF.....EKGILDKERAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1957	81.6	533	2 A37180	chromogranin/secr
2	196	8.2	2139	2 T18296	myosin heavy chain
3	191	8.0	2454	1 QRMSP1	microtubule-associ
4	186.5	7.8	2116	2 A26855	myosin heavy chain
5	184	7.7	665	2 B71609	hypothetical prote
6	183.5	7.6	1790	2 S67593	transport protein
7	180.5	7.5	2364	2 A56577	microtubule-associ
8	179.5	7.5	1526	2 A45605	mature-parasite-in
9	179	7.5	853	2 T51505	hypothetical prote
10	179	7.5	1005	2 A64465	hypothetical prote
11	179	7.5	1558	2 B71603	RESA-H3 antigen p
12	179	7.5	2269	2 T28677	rhostry protein -
13	178.5	7.4	1392	2 A43336	microtubule-vesicl
14	178.5	7.4	1427	2 S22695	restin - human
15	177	7.4	1871	2 D96796	probable heat shoc
16	172.5	7.2	1134	2 A60234	Iga Fc receptor pr
17	172.5	7.2	1164	1 FCSOAG	Iga Fc receptor pr
18	172.5	7.2	1979	2 C71622	hypothetical prote
19	171.5	7.1	3488	2 T34418	hypothetical prote
20	170	7.1	1804	2 T34518	nestin golden ha
21	169	7.0	1272	2 C90593	hypothetical prote
22	168.5	7.0	746	2 T47237	myosin II heavy ch
23	168.5	7.0	926	2 T24923	hypothetical prote
24	168	7.0	2401	2 T28676	rhostry protein -
25	167	7.0	976	2 T01553	hypothetical prote
26	166.5	6.9	1875	2 S38173	myosin-like protei
27	166.5	6.9	5327	2 T13564	microtubule-associ
28	166	6.9	991	2 H86168	hypothetical prote
29	166	6.9	1192	2 A71623	probable secreted

30	166	6.9	2663	1 S28261	centromere protein
31	164.5	6.9	719	2 A81358	hypothetical prote
32	164.5	6.9	911	2 S51441	hypothetical prote
33	164.5	6.9	1127	2 T28317	ORF MSV156 hypothe
34	164.5	6.9	1269	2 F84730	probable myosin he
35	164	6.8	1909	2 A45592	liver stage anti
36	163	6.8	1252	2 B42771	reticulocyte-bind
37	163	6.8	2510	2 T28160	hypothetical prote
38	162.5	6.8	1738	2 T14967	interaptin - slime
39	162.5	6.8	2385	2 A32491	myosin heavy chain
40	162.5	6.8	2411	2 B32491	myosin heavy chain
41	161.5	6.7	1302	1 JC6009	surface-located me
42	161.5	6.7	1992	1 S02771	myosin heavy chain
43	161	6.7	852	2 D72230	conserved hypothet
44	160	6.7	1354	2 S74244	serine/threonine-s
45	159.5	6.6	821	2 S67087	hypothetical prote

ALIGNMENTS

RESULT 1

A37180

Chromogranin/secretogranin-like vesicle protein precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 30-Sep-1993

C;Accession: A37180

R;Otterger, H.P.; Battenberg, E.F.; Tsou, A.P.; Bloom, F.E.; Sutcliffe, J.G.

J. Neurosci. 10, 3135-3147, 1990

A;Title: IB1075: a brain- and pituitary-specific mRNA that encodes a novel chromogranin/

A;Reference number: A37180; MUID:90376160; PMID:2204688

A;Accession: A37180

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-533 <OTT>

Query Match 81.6%; Score 1957; DB 2; Length 533;

Best Local Similarity 86.7%; Pred. No. 3.5e-87;

Matches 383; Conservative 21; Mismatches 38; Indels 0; Gaps 0;

QY	16	PTQAFPKPGSQKSLHNRSLSAERPLNEQIAEAEADKIKKTPPENKPGQSNVSVFVNL	75
DB	63	PNSFPKPGSQKSLHNRSLSAERPLNEQIAEAEADKIKKTPSESKPSERFSSVDNL	122
QY	76	NLLKATTEKTEKTERQSRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSLDHFODDP	135
DB	123	NLLKATTEKTEKTERQSRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSLDHFODDP	182
QY	136	DGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAE	195
DB	183	DGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAE	242
QY	196	VLOKLISKANNYEEDPNKPTSWENOAGKIPKVTPTMAIQDGLAKGENDETIVSNLTLL	255
DB	243	ALQKLISKANNYEEDPNKPTSWENOAGKIPKVTPTMAIQDGLAKGENDETIVSNLTLL	302
QY	256	TNGLERTKTYSEDNEFEELQYFNFYALLKSIDSEKEAKEKETLITIMTKLIDFVQNMVK	315
DB	303	SNGLERTKTYSEDNEFEELQYFNFYALLKSIDSEKEAKEKETLITIMTKLIDFVQNMVK	362
QY	316	YGTISPVEGVSVLENIDEMIALQTKNLEQATDNISKLPAPSEKSHETDSTKEAAK	375
DB	363	YGTISPVEGVSVLENIDEMIALQTKNLEQATDNISKLPAPSEKSHETDSTKEAAK	422
QY	376	MEKEYGSLKSTKDDNSNPGGKTDEPKGTAEYLAIRKNIEWLKKHKDKGNKEDYDLK	435
DB	423	MEKEYGSLKSTKDDNSNPGGKTDEPKGTAEYLAIRKNIEWLKKHKDKGNKEDYDLK	482
QY	436	MRDFINQADAYVEKGILDKKEE	457
DB	483	MRDFINQADAYVEKGILIRKK	504

RESULT 2

T18296
 myosin heavy chain - Entamoeba histolytica
 C:Species: Entamoeba histolytica
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
 C:Accession: T18296
 R:Guillen, N.
 submitted to the EMBL Data Library, February 1997
 A:Reference number: 218865
 A:Accession: T18296
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2139 <GUI>
 A:Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
 C:Genetics:
 A:Gene: mhca
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 8.2%; Score 196; DB 2; Length 2139;
 Best Local Similarity 21.0%; Pred. No. 0.1;
 Matches 107; Conservative 108; Mismatches 191; Indels 104; Gaps 23;

Qy	20	FPKGGSDKSLHRELNSAERPLNEQIAAEEDKIKTYPPE-----KPGOSNYSFV	72
Db	859	FEPEGKKOKIED-----LKKLAE-----EIKKRAENALAGATATGEBLEAKIQ	906
Qy	73	D-----NLNLKAITKEKEKIEKQIRSSPLDNKLNVEDVD-	109
Db	907	DLEDKISELSKLSAELDQKELNLKLENLEEDKEELKETTIDNLKGLDKSLKAGEDLEV	966
Qy	110	-----STNKRKLIDYDSTKGLDHKFDQDDPLHQLDGTPLTADIVHKLARIY	160
Db	967	EITELNSQINTLN-ATVDKDKTIAEMQESIDEKEDEITKLKGDIKLLEEKDDL-----	1020
Qy	161	EENDR-----AVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLSKANNVEEDPNKPT	216
Db	1021	-EQDRADVSATKDDTAKKLNKITI-----ECEDAKDEIAKLQLELEDEB--NKNKDLTNEL	1073
Qy	217	SWTNOAGKIPKVTPTMAAI-----QDGLAG--ENDETVNTLTTLNG--LERRTKYS	267
Db	1074	QQTQLKLGTEKSLAAQVAATKASDRDITLSONLENEKLTITKNTTKTKADLEKKISGLK	1133
Qy	268	EDNFELOVFPNFFYALLKSID-----SEKEAKEKTLITIMKTLIDFV--KMWVYGTITSP	321
Db	1134	QD-YEDLEDDKN-----KIEGLRNRQKIKELDDEITKGADVSQYLQOKKEEY-----	1181
Qy	322	EBGVSYLENLDEMIALQTKNLEKNATONISKLPAPSEKSHETDSTKEAAK-----ME	377
Db	1182	ESQIAMQEEKKAIQNDVKNK-EKTIKE--KELEIQSLQEKLDTEVEKEDEAKKKKEIE	1238
Qy	378	KRYGSLKSTKDDNSNPGKTDPEKGTAYLEATRKNIEMLWKHKDKGNKEDYDL-SKM	436
Db	1239	KEMKALQEE--KENVSSKNSTKDKKKLLEDNLKDTQKKLLDMDTADNEKLKAKAKDLEAQL	1297
Qy	437	RFINKQADAYVEKGILDEAEATKRIYS	466
Db	1298	NEVDQNHKAVADAELLNKKKAQSDKELNS	1327

RESULT 3
 QRMSP1
 microtubule-associated protein MAP1B - mouse
 N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
 C:Species: Mus musculus (house mouse)
 C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
 C:Accession: S07549; S44387; A33645
 R:Noble, M.; Lewis, S.A.; Cowan, N.J.
 J. Cell Biol. 109, 3367-3376, 1989
 A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
 A:Reference number: A33645; MUID:90094539; PMID:2480963
 A:Accession: S07549
 A:Molecule type: mRNA

A:Residues: 1-2464 <NOB>
 A:Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000
 R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
 Arch. Biochem. Biophys. 310, 428-432, 1994
 A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
 A:Reference number: S44387; MUID:94234720; PMID:8179328
 A:Accession: S44387
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 653-663, 'IC' <SAN>
 C:Superfamily: microtubule-associated protein MAP1B
 C:Keywords: microtubule binding; phosphoprotein; tandem repeat
 F:589-786/Domain: microtubule binding #status experimental <MTB>
 F:589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-694;
 R-K-E/D-X)
 F:1861-2064/Region: 17-residue repeats
 F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: phosphate (Thr) (co
 F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Tyr) (covalent) #status predicted
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 8.0%; Score 191; DB 1; Length 2464;
 Best Local Similarity 21.3%; Pred. No. 0.21;
 Matches 118; Conservative 85; Mismatches 186; Indels 166; Gaps 24;

Qy	22	KPGGSDKSLHRELNSAERPLNEQIAEA-----EEDKIKTYP--PENKPGQSNYSFVD	73
Db	550	KPVASKSVRKESKEETPEVTKTSQVEKTPKVESKEVVLVKDKKVKTESKP-----	600
Qy	74	NLNLKAITKEKIEKQIRSSPLDNKLNVEDVDSTKRLIDDDYSTKGLDHKFD	133
Db	601	-----SVTEKVSSEKESQSPVKAQVAKQATKSPKVTQKVVVKKEIKTK--LEEKKEE	652
Qy	134	DP--DGLHOLDGTP-----TAEDIVHKAIAARIYENDRAVFDKIVSKLNLGLITESQ	185
Db	653	KPKKEVVKKEDTLPKDKDEPRKEEVKKEIKKEIKK--ERKELKKEVKK-----ETP	703
Qy	186	AHTLEDEVAEVLQKLSKANNVEEDPNK-----PTSWTENQAGKIPE-----	228
Db	704	LKDAKKEVKEEKEVKKE-----EKEPKKEIKKIKSKDKIKKSTPQSDTKKPSALKPKVAKK	759
Qy	229	-----KVTPTMAA--IQD-GLAK-----GENDETVENTL-----	253
Db	760	BESTKKEPLAAGLKDKGKVKVIKKEGKTTTAAATAVGTATTAATAVVAAGIAGSPVKE	819
Qy	254	-----TLTNGLERTKTYSNEDNFELOVFPNFFYALLKSIDSEKAKE-----KET	298
Db	820	LEARSLSMSPEDLTQKFEELKAEIDVAKDIKQLELIEDEKLEKETQPGAEVVIQKET	879
Qy	299	LITIMKTLIDFVRMWVYGTITSPSEGVSYLENLDEMIALQTKNKL--EKNATONISKL-	354
Db	880	EVS-----KGSAESPDGEGITTTTGEGE--CEQTPPELEPVEKQGVDDIEKPE	924
Qy	355	-----FPAPSEKSHETDSTKEAAKQKESGLSKDSTKDDNSNPGKTDPEKGTAYL	409
Db	925	DEGAGFESSETGYEKAETETAEAEPEED--GEDNASASAKSHSTEDDESAAKADVHL	983
Qy	410	PAIRKNI-----EWLKK-----HDKNGKNEDY-----DLKSKWRDFIN	441
Db	984	KEKESVSVGGDRAEEDMDVLEKGEAQSEEGEEDKAEDAREGVEPDKTEADYVM	1043
Qy	442	KQADAYVEKGILDK	456
Db	1044	AVADKAAEAGVTEEQ	1058

RESULT 4
 A26655
 myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Dictyostelium discoideum
 C>Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002
 A:Accession: A26655; A24728; S00250
 R:Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum
A:Reference number: A26655; MUID:87092266; PMID:3540939
A:Accession: A26655
A:Molecule type: DNA
A:Residues: 1-2116 <WAR>
A:Cross-references: GB:M14628; GB:M11938; NID:g167834; PIDN:AAA33227.1; PID:g167835
R:DeLozanne, A.; Lewis, M.; Spudich, J.A.; Levinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A:Reference number: A24728; MUID:86016788; PMID:3901008
A:Accession: A24728
A:Molecule type: mRNA
A:Residues: 2035-2116
R:Wagile, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium A:Reference number: S00250; MUID:88112226; PMID:2828113
A:Accession: S00250
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1734-1893 <WAG>
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phosphatase
F:1-818/Domain: globular head <HD>
F:89-747/Domain: myosin motor domain homology <MMOT>
F:179-186/Region: nucleotide-binding motif A (P-loop)
F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 7.8%; Score 186.5; DB 2; Length 2116;
Best Local Similarity 21.6%; Pred. No. 0.29;
Matches 108; Conservative 86; Mismatches 205; Indels 101; Gaps 17;

Qy 26 SODKS--LHNELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSFVDNLNLKAITTE 83
Db 1023 TKDSKELLRQKKLEELKQVQEAALAEATAKLAQAANKLQGEYTEL-NEKFNSEVTA 1081

Qy 84 KEIEKERQSRSS--PLDNKLVNEDVDSTKRNKLIIDYDSTKSLGDLHKFDQDGLHQL 141
Db 1082 RSNVEKSKTLESQVAVNNELDEE---KKNR---DALEKKKALDAMLEEMKQDLEST 1134

Qy 142 DGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGLITESQAHTLDEVAEVLQKI 201
Db 1135 GGEKSKSLYD-----KYQESDMALRNQISLQSTIAKLEKIKTLGEGVARLQGELE 1188

Qy 202 SKE--ANNYEEDPNKPTSWTENQAGKIPKVT-----PMAAIDQGLAKGEN 245
Db 1189 AEQLAKSNVEKQKKVVELDLDEKSAQLAEETAQAQALDKLKKLEQBLSEVQTOLSEANN 1248

Qy 246 D-----ETVSNLTITLNGLERKTKTYSEDN-----FEELQYFPNFYALLKSIDSE 290
Db 1249 KVNNSDSTNKHLETSFNNLKLELEAEQAKQALEKKRIGLSELSKH-----VNEQLEEE 1302

Qy 291 KEAEK-----ETLITIMKTLDIFVMMVKGTITSPBEGSVYLE----- 329
Db 1303 KQKQESNEKRVLDLEKEVSELQDTEEVASKAVTEAKNKKSELSDELIRKQYADVSSR 1362

Qy 330 --NLDEMTALQTN-KLEKNATDNISKLFPAPSEKSHETD-----STKEAAKMEKY 380
Db 1363 DKSVEQLTKLQAKNEELNRTAEAGQDLRAERKKAEPDLEAVKNLEETAKKVAKE 1422

Qy 381 GSLKSDTKDDNSNPGKTDPEPKGTEAYLEAIRKNIEWLKHKKNKEDYDLGSKMRDFI 440
Db 1423 KAMKKAETDYRSTKSELDDAKNVSEQVQIKRLNEE-----LSLSRVL 1467

Qy 441 NKQADAYVEKGILDKAEBA 460
Db 1468 -EADERCNSAIKAKKTAES 1486

RESULT 5
B71609
hypotheical protein PFB0680w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71609
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Atavind, L.; Koonin, E.V.;
Perteau, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: B71609
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-665 <GAR>
A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AAC71925.1; PID:g3845245
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0680w

Query Match 7.7%; Score 184; DB 2; Length 665;
Best Local Similarity 22.0%; Pred. No. 0.089;
Matches 104; Conservative 76; Mismatches 182; Indels 110; Gaps 17;

Qy 25 GSODKSLHNELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSFVDNLNLKAITTE 84
Db 160 GKQDISNSAE--NKQDVKGVELEKKEKESISDDHKVEENKKS-----DD 205

Qy 85 EKIEKERQSRSSPLDNKLN---VEDVDSTKRNKLIIDYDSTKSLGDLHKFDQDGLHQL 140
Db 206 HKVEENKKSDDHKVEENKKSDDHKIEVKKVEEHEDEEDKKEKSENKKNKDKEND 265

Qy 141 LDGTPLTAEDIVHKIAARIYEENDRAVDFKIVSKLNLGLITESQAHTLDEVAEVLQKL 200
Db 266 EDNDEISDEVEDDDVEEDKNENDIDDDK-----KETDKTHLEEEENEIEKE 314

Qy 201 IS-KEANNYEEDPNKPTSW-TENQAGKIPKVTPEMAAIDQGLAKGENDETIVSNLTITLNG 258
Db 315 FSKKQKGNKNDTKKESKOTEKESKDIEK-----EKSKEKEKS----- 356

Qy 259 LERRTKTYSBDFEELQYFPNFYALLKSIDSEKAEKETLITIMKTLDIFVMMVKYGT 318
Db 357 -KQKEKQKQKEK-----KSDIEKE-KEKQDIEKESK-DTAKEKEKDKD 402

Qy 319 ISPEGVSVLENLDEMTALQTNKLEKNATDNISKLFPAPSEKSH-----E 364
Db 403 IEKEKS-----KMEKLNKQNDKEDKDDNEKK-----KNDKQDIHDDNDENDMBEIE 451

Qy 365 ETSTKEAAKME-----KEYGSLKDSPTK-DNNSNPGKTDPEPKGTEAYLEAI 412
Db 452 ENDEDEDEDMENKKKKGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 511

Qy 413 RKNIEMLKHHDKGNKEDYDLGSKMRDF-----INKQADAYVEKGILD 454
Db 512 NENNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 563

RESULT 6
S67593
transport protein USO1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D2552; protein YDL058w
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C:Accession: S67593; A38455; S30782
R:Bloeker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67587
A:Accession: S67593
A:Molecule type: DNA
A:Residues: 1-1790 <BLO>
A:Cross-references: EMBL:274106; NID:g1431058; PID:e253003; PID:g1431059; MIPS:YDL058w
A:Experimental source: strain S288C
R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
J. Cell Biol. 113, 245-260, 1991
A:Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transport
A:Reference number: A38455; MUID:91185402; PMID:2010462

A;Accession: A38455
A;Molecule type: DNA

A;Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>

A;Cross-references: GB:X54378; NID:94777; PIDN:CAA38253.1; PID:94778

A;Note: the authors translated the codon ACT for residue 768 as Ile

R;Hottel, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.

submitted to the EMBL Data Library, February 1993

A;Description: An integrin analogue in Saccharomyces cerevisiae.

A;Reference number: S30782

A;Accession: S30782

A;Molecule type: DNA

A;Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580, 'S'

A;Cross-references: EMBL:L03188

C;Genetics:

A;Gene: SGD:USO1; INT1

A;Cross-references: SGD:S0002216; MIPS:YDL058w

A;Map position: 4L

C;Keywords: coiled coil; transmembrane protein

F;326-342/Domain: transmembrane #status predicted <TM1>

F;394-410/Domain: transmembrane #status predicted <TM2>

F;617-633/Domain: transmembrane #status predicted <TM3>

Query Match 7.6%; Score 183.5; DB 2; Length 1790;

Best Local Similarity 20.3%; Pred. No. 0.33;

Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;

Qy 29 KSLHN--RELSAERPLNEQIAEAEEDKIKTYPPENKPGQSNYSFVDMNLNLLKAI---T 82

Db 965 KSLANNYKMQAE--NESLIKAVEE-----SKNESSIQLSNLQNKIDMSQ 1008

Qy 83 EKEIEKERQIRSPDLNKLNVEDVSTNKRKLI-----DDYSTKSGLDHFKPD--- 133

Db 1009 EKENFOIERGSIENIKNIEQLKKTITSDLEQTEIEIISKSDSSKDEYESQISLLKELTATT 1068

Qy 134 -DPDGLHOLDGTPLEADIVHKIARIYEENDRAVDFKIVSKLNLGLITESQAH----- 187

Db 1069 ANDENVNKISLTKTRBELEALAA--YXNLKNELETKLETSEKALKEVNEHKLKEEK 1126

Qy 188 -TLEDEVAEVLQKLISKEAN-----NVEED-PNKPTSWTNOAGKIPEK 229

Db 1127 IQLEKEATETKQQLNSLRANLESLEKEHDLAAQLKYEEQIANKERYNE-EISQLNDE 1185

Qy 230 VTPWAAIQDGLAKGENDETIVNTLTNGLERRT-----KTYSEDN-- 270

Db 1186 ITTQQENESIKK-KNDELEGEVAMKSTSEQSNLKKSEIDALNLQIKELKKNETNEA 1244

Qy 271 -----PELOVFPNF-----YALLKSIDSEKAK 294

Db 1245 SLESIKSVSEETVKIKELQDECNFKKEVSELEDKLSKASEDKNSKYLELQK-ESEKIK 1303

Qy 295 EKETLITIMKTLDIFVQMVK----- 315

Db 1304 ELDAKTTTELKIQLEKITNLKAKESSELSRLKTSSEERKNAEBQLEKLNKIQIKNQ 1363

Qy 316 -----YGTISPSE--EGSVYLENDEMIALQTKNKLKNATNTSKLPPAPSEK 361

Db 1364 APFKERKLLNEGSSITITQYSEKINTLE--DELIRLQNELKAKIEDNTRSELEKVSLS 1421

Qy 362 SHEETDSTKEAAKMEKEYGSLKDS--TKDNSNPGKTDPE-----KKGTEAVLEAIRK 414

Db 1422 NDLELLEKQNTIKSLQDEILSYKDKITRNBEKLLSIERNKRDLESLEKQIRAAQESKAK 1481

Qy 415 NIEWLKHKDKGNKEDYLSKMRDFINKQADAVVEKGILD-KEEAAIKR 463

Db 1482 VERGLKLEBESSKAELEKSEMM-KLESTIESNETELKSSMETIRK 1530

RESULT 7

A56577

A;Title: microtubule-associated protein MAP 1B - rat (fragment)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997

C;Accession: A56577

R;Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.

Eur. J. Cell Biol. 57, 66-74, 1992

A;Title: Identification of two distinct microtubule binding domains on recombinant rat M

A;Reference number: A56577; MUID:92347374; PMID:1639092

A;Accession: A56577

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-2364 <ZAU>

A;Cross-references: GB:X60550

A;Experimental source: brain

A;Note: nucleotide sequence not given; conceptual translation not complete

C;Superfamily: microtubule-associated protein MAP1B

Query Match 7.5%; Score 180.5; DB 2; Length 2364;

Best Local Similarity 22.7%; Pred. No. 0.65;

Matches 120; Conservative 77; Mismatches 200; Indels 131; Gaps 22;

Qy 22 KPGSQDKSLHNRSLAERPLNEQIAEAEEDKIK-KTYPPENKPGQSNYSFVDMNLNLLKA 80

Db 454 KPLSSKSVRKESKEAPEATKASQVEKTPKVESKEKIVVKDKPGK-----VESKPS 505

Qy 81 ITEKEKIEKERQIRSPDLNKLNVEDVSTNKRKLIIDDYSTKSGLDHFKFODDP--DGL 138

Db 506 VTEKEVPSKEQSPVKAFAEAAKAAATESKPKVTKYKVKVKEIKTKP--BEKKEKPKKEVA 563

Qy 139 HQLDGTPLTARDIVHKIARIYEENDRAVDFKIVSKLNLGLITESQAH--LEDEVAEVL 197

Db 564 KKEDKTPPKDEKPKK-----BEAKKEIKKEKEKELKKEVKETPLKDAKKEV- 615

Qy 198 QKLISKEANNVEEDPNK-----PTSWTENQAG-----KIPEKVTMAA- 235

Db 616 KKDEKKEVKEKEPKKEIKKIDKIKSTPLSDTKPKPAALPKPVAKKEEPTKKEPIAAG 675

Qy 236 -IQD-GLAK-----GENDETVNTL-----TLTNGLERRTK 264

Db 676 KLKDKGKVKYKKGKTEAAATAVGTAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 735

Qy 265 TYSNDNFEELOQYFNFVALLKSIDSEKAEK-----KETLITIMKTLDIFVQMVV 314

Db 736 DFELKAAEEDVAKIDKIPQLELIEDEKLEKTEGEAVVIOKETEVS----- 782

Qy 315 KYGTISPGEVSVYLENDEMIALQTKNKL-----EKNATDNISKL-----FPAPSKSHEE 365

Db 783 KGSAESPDGEGITTTTEGE--CEQTPPELEPVKQGVDDIEKFEDEGAGFESSEAGDYE 840

Qy 366 TDSKTEAAKMEKEYGSLKOSTKDDNSNPGKTDPEKPKTEA--YLEAIRKNI----- 416

Db 841 EKAETEEAEEPDG-----EDNVSGSASKHSPTDEETIAKAEADVHIKEKRESVASGDDRA 897

Qy 417 -----EWLKKHKDKGNKEDYLSKMRDFINKQADAVVEKGILDKEEAE 459

Db 898 EEDMDALEKGEAQSEEEGEE-----DKAEDAREDEHDPDKTEAE 940

RESULT 8

A45605

A;Title: mature-parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium

C;Species: Plasmodium falciparum

C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C;Accession: A45605; A54517

R;Coppel, R.L.

Mol. Biochem. Parasitol. 50, 335-347, 1992

A;Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human ery

A;Reference number: A45605; MUID:92158014; PMID:1741020

A;Accession: A45605

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1526 <COP>

A;Experimental source: Papua New Guinean isolate FC27

A;Note: sequence extracted from NCBI backbone (NCBIN:83648, NCBI:83656)

R;Coppel, R.L.; Culvenor, J.G.; Bianco, A.E.; Crewther, P.E.; Stahl, H.D.; Brown, G.V.; /

Mol. Biochem. Parasitol. 20, 265-277, 1986

A;Title: Variable antigen associated with the surface of erythrocytes infected with matsu

A:Reference number: A54517; MUID:87014571, PMID:35318494
A:Accession: A54517
A:Status: preliminary
A:Molecule type: mRNA
A:Cross-residues: 222-443, 'K', 445 <CO2>
A:Cross-references: GK:M15319; NID:g160060; PID:g552170
C:Superfamily: king-infected erythrocyte surface antigen
C:Keywords: surface antigen; tandem repeat

Query Match	7.5%;	Score 179.5;	DB 2;	Length 1526;
Best Local Similarity	18.7%;	Pred. No. 0.42;		
Matches 107;	Conservative 101;	Mismatches 182;	Indels 181;	Gaps 24;

Qy	25	GSODKSLHNRELS--AERPLNEQIAEAEEDKIKKTYPPENKPGQN-----	68
		: : : : : : : : : :	
Db	400	GASINSEDPKLTQEENGTKESSEETKDDK-----PEENEKKADNKKKKKKKKSPFQM	454

Qy 69 --YSFVDNLNL-----LKAITEKEIEKERQSIRSPDLNKLNVEDVDSTQNRKLIDDYD 121
Db 455 LGCNPLCNKNNIETDDEEETLVVKDDAKKHKFLREA-----NTSKNDNEKKDKULGEGD 508

Qy	122	-----STKSLDHKFQDDPGL-----	-----HQLDGTPL	146
			:	:
Db	509	KEDVKEKNDKQKVLGEGDKEDVKEKNDKQKVLGEGDKEDVKEKNDKQKVLGSEK	568	

```
Qy      147 TAEDIVHKAARI-----YEENDRAVFDKIVSKULNLGILITESQAHTLEDEV-- 193
        |::| |: |: :||| ||: |::| ::| ::| ::| ::| ::| ::|
Db      569 TORIKERKEVEIRVKKKKKVKKGIGKENDEGNDRKVGGPRIIEEVKEETKKOVGDGIKE 628
```

QY	194	-----AEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPKVT-----MAAI	236
		:::	
Dp	629	NDTEGNDKVKGPETTEVKFEIKKOVEDCIKENDETEGNDKVKGPETTEVKFEIKKOV	688

Qy	237 QDGLAKG--ENDETVSNTLTLTNGIERRTK-----TYSDNF-----EELQY 276
Dh	EAGIKENITECNDKVKGDRTITREVFETKKOVERECIKENDTESCKVTCBITTPPVV- 247

Qy 277 FPNFYALKSIDSEKAEKETLITI-----MKTLIDFQWVKVGTISPEGV--SYL 328
| | : : | | : :
Db -----XPIEVAECVECHVENTIETIVNUTCAEVITIEBKVYUW VTECVTKEVNTIM 306

QY 329 ENLD-----EMIALQTKNLEKNATDNISKLPAPSEK-----SHEETDSTKEEA 373

Qy 374 AKMEKEYSLKDSWKDDNSNPGGKTDEPKCKTEAYLEAIRKNI EWLKKHKDKGNKEDYDL 433

Q7 434 SK--MRDFINKQADAYVEKGILDKEEAIAIK 462
 _ | :: : | :: | : | :: | : |

RESULT 9

hypothetical protein F5E19_70 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000

R/Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabatake, Y. *Protein Data Bank*. 1995
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394

A;Accession: J31505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-853 <SAT>

A;Cross-references: EMBL:AL391114/
A;Experimental source: cultivar Columbia; BAC clone F5E19
C;Genetics:
A;Map position: 5

A;introns: 6/2; 79/3
A;Note: FSE19_70

Query Match	7.5%	Score 179;	DB 2;	Length 853;
Best Local Similarity	20.9%	Pred. No. 0.21;		
Matches 102;	Conservative	81;	Mismatches 206;	Indels 100;
Gaps 19				

QY	16	P	I	O	A	P	F	F	E	G	S	D	K	S	L	N	R	E	L	S	A	R	-----	P	L	N	E	O	I	A	E	A	E	D	K	I	K	T	T	P	E	N	-----	62										
DB	45	P	S	T	T	P	H	S	R	L	S	D	R	S	S	N	K	S	S	V	E	R	R	S	P	K	L	P	T	P	E	K	S	O	A	R	V	A	A	V	K	G	T	E	S	O	T	T	L	S	O	I	K	104

QY 63 ----KPGQSNYSFVDNLLNLLKATTEKEKTEKERQSIRSSPLDNKUNVEDVDSTKNRKLIDD 119

DB 105 EDIKKANERISSLEK-DRAKALDELQAKKEAEQV-TLKLDLDAKQ--KHVEENSEIEK 160

```

QY 120 YOSTKSLDHKFFQDDPGCL-----HQLDGTLPTLAEDIHVHKIAARTYEENDRAVFD 169
: : : : :
Db 161 FQAVEAGIE-AVONNEESLKKELETVKQHASDAAALVA---VQEELEKINEEL-AAAF 215

```

QY 170 KIVSKLNLGLINESQAHTLEDEVAEVLQKLSKEANNVEEDPNKPTSWTENQAKIPEK 229
DB 216 -AKSKAL-----SOA-----EDASKTAEIHAEKVDILSSE 244	

QY		230	VTPMAAIQDGLAKGENDETFSNTLTTLTGLERRRTKTYSDNFEEELQYFFNYALLKSIDS	289
	:	:	:	:
DB		245	LITFLKALID-----STREKTATISDNEMVAKIRDEIVLKKDRIRS	283

QY	290	----	E	K	A	K	E	K	E	T	L	I	N	K	T	L	I	D	P	V	K	M	V	K	Y	G	T	I	S	P	E	G	V	S	Y	L	N	D	E	M	I	A	L	O	T	K	N	K	L	E	K		345
DH	284	A	R	G	F	F	A	F	V	K	E	M	I	V	E	K	T	N	V	I	D	I	P	A	A	K	A	E	S	N	A	H	S	I	S	N	E	W	O	S	K	A	P	E	I	P	E	O	----	341			

QY . 346 NAT--DNISKLFPAPSEKSHETDST--KEBAKMEKEYGSLKJSTKDDNSNPGKTD 399

Qy 400 EPKGKTEAYLEARKNIEWLKXHDKGKEDYDLSKMRDFINKQADAVVEKGILDKEEAE 459

Qy	460	AIKRIYSL	468
		:	
	461	WZJWZJ	469

RESULT 10

hypothetical protein MJ1322 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Gerson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus*
A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: A04783
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1005 <BUL>

A:Map position: REV1273394-1270377
 C:Superfamily: hypothetical protein MJ1322
 A:Cross-references: GB:067572; GB:L1/111; NID:G1591958; F1DN:MA699331.1; PID:1

Query Match	7.5%	Score 179;	DB 2;	Length 1005;
Best Local Similarity	22.3%	Pred. No. 0.26;		
Matches 116;	Conservative	80;	Mismatches 164;	Indels 160;
				Gaps 27

Qy	52	DKIKTYPPENKPGQSNYSFVNINLLKA-ITEKEKIBERQSIIRSSPLDNKLN-VEDVD	109
Dh	159	DPEKCY--OMGEIYKVEYKELTETGEENYKENYEKE-----LKNKMSQIER--	205

QY 110 STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169

RESULT 9
T51505
hypothetical protein F5E19_70 - *Arabidopsis thaliana*

C/Species: *Arabiopsis thaliana* (mouse-ear cress)
C/Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C/Accession: T51505
R/Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew

submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51505
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-853 <SAT>
A;Cross-references: EMBL:AL391147
A;Experimental source: cultivar Columbia. BAC clone F5E19

C:Genetics:
A:Map position: 5
A: Introns: 6/2: 79/3
B: Note: 55F19.70

```

Db 206 --KNKLMKLNK- IKKPFEDIEKLFNENKLIYKFKINL-----EKKRALEL 257
QY 170 KIVS-KLNLGLITISQAH-----TLEDEVAEVLQKLSKEANNYEEDPNK 214
Db 258 KNOELKILEYDLNTVVEARETLNRHKBREYKYSIVDBIRKIESRL--RELKSHVEDYK 315
QY 215 PLSWTENQAGKPEKVTWMAAIOGLAGENDETVSNTLTNTGLERRTKTYSNDEEEL 274
Db 316 LTKQLEIKGDI-EK-----LKEPINKSKYRDDIDNLDLTKNI-----KDEIERV 360
QY 275 QYFPNFYALLSIDSEKAKK-----ETLITIMKTLDIFVOMVYKGTISPE 322
Db 361 ETIKDLLEELNLENEIEKIKYRICECKEYVEKYLEBEKAVEYNKLTLEYITL--- 417
QY 323 EGVSYLENDEMIALQTNKLEKNATD---NISKLPAPS-----EKSHEEDTSTK-- 370
Db 418 -----LOEKKSEKINDLETRINKLLEETKNDIENSLKETEKKV 463
QY 371 -----EEAAKMEKVGSLKOSTK-----DDNSNPGGK-----TDEPK- 405
Db 464 LENLQKEKIELNKLGEINSEIKLKKILDDELKEVEGCKPCKTPIDENKMWELNHQKT 523
QY 406 E--AYLE-----AIRKNIEWLKKG-----DKGKNKEDYDL----- 433
Db 524 QLNKKYTELEBINKIREIEKDKKEIDKEENLTKLTKLYLEKQSQIEELEKLKNY 583
QY 434 SKMRDFINKQADAYVEKG-----TLD--KEEAIAIKRYS 466
Db 584 KEQJDEINKKISNVINGKPVDEILEDIKSQLNKFKNFYN 623

RESULT 11
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71603
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Peretea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: B71603
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1558 <GAR>
A:Cross-references: GB:AE001424; GB:AE001362; NID:G3845307; PIDN:AACT1972.1; PID:G384530
A:Experimental source: clone 307
C:Genetics:
A:Gene: PFB0915w

```

```

Query Match 7.5%; Score 179; DB 2; Length 1558;
Best Local Similarity 21.5%; Pred. No. 0.45;
Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;

QY 27 QDKSLNRELASB--RPLNEQIAEAEEDKIKTYPENKPGQSN-----YSFVDNL-- 75
Db 725 EENAVESNENVAENLEKLNATVNTVLDKVEETVEISGESLENEMDKAPFSEIFDNVKG 784
QY 76 ---NLK-----ALTEKEKIEKQSRSSPLDNKLN-----VEDVDSTK-- 112
Db 785 IQENLLTGMFRSTETISIVIOSEKVDL--NENVVSSILDNENKMGKLNKLENIESTEGV 843
QY 113 -----NRKLDDYD-----STKSGLDHKFQDDPDGLHQLDGPPLTAEDI- 151
Db 844 QETVTEHVEQNVVDVDPAMKDQFLGILNEAGGLKEMFNLED-VFKSESVDIVIEIK 902
QY 152 -----VHKIAARIYEENDRAVDFKIVSKLNLGLITISQAHTEDEVAEVLQKLSKE 204
Db 903 DEPVQKEVSEKTVTSIIEENIVD-----VLEEKEDLTDKMDIAVESIEIS 951
QY 205 ANNYEEDPNKPTSWTENQAGKIPEKVTWMAAIOGLAGENDETVSNTLTNTGLERRTK 264

```

```

Db 952 SDSKEE-----TES-----TKOKEKOVSLVVEEQDNDMDSEKVLKLMEEELMK 999
QY 265 TYSEDN-----FEELQYFPNFYA-LLKSIDSEKAK-----EKETLITIMK 304
Db 1000 DAVEINDITSKLIEETQELNEVEADLIKMEKLEKALSSEDSKEIIDAKODTLEKVIK 1059
QY 305 -----TLIDFVOMVYKGTISPE--EGVSYLENDEMIALQTK--NKLKKNATONISKL 354
Db 1060 EEHDITITLDEV---VELKDVBEEDKIEKVDLKEEDILKEVKEIKELSEBILEDYKEL 1116
QY 355 FPAPSSKSEHEEDTSTKEEAAMEKVGSLKOSTKDDNSNPGKTDPEKCKTAYLEAIRK 414
Db 1117 KTIETDILKEKIEKDHPEKEEAEETKOLEADILKEVSVLEVEEKKLESEVHE--LKE 1175
QY 415 NIEWLKHKDK--KGNKEDYDLKMRDFINKQADAY---VEKGILDK-----EKA 458
Db 1176 EVEHIISGDAHIKGLEED-DLEEVDLKGSLDMLKGMELGDMDKESLVEDVTAKLGERV 1234
QY 459 EAIKRIYSS 467
Db 1235 ESLKQVLSS 1243

RESULT 12
T28677
rhoptry protein - plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: Z20508; MUID:95021522; PMID:7935623
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEE>
A:Cross-references: EMBL:L27838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 7.5%; Score 179; DB 2; Length 2269;
Best Local Similarity 21.1%; Pred. No. 0.73; Mismatches 85; Indels 144; Gaps 23;
Matches 113; Conservative 85;

QY 13 LVLPQIAPPKGGSQDKSLNRELASERPLNEQIAEAEEDKIKTYPENKPGQSNYSFV 72
Db 7 LILPLNQFKGLNESMIKLNKSGILRKVTISNQI---KNKLVNSTYPEGREG-----FT 57
QY 73 DNALN-----LKAITEKEK-----TEKQSRSSPLDN-----KLN- 104
Db 58 SSLELAKSWKTKLETITELTKNEETVRLEKEIRELFKYLDEAEARKYLEGLKLELNK 117
QY 105 -----VEDVDST-----KNRKLIDD-----YDSTKSLGDH 129
Db 118 KIKDIIAKIEVKNVTVELKKEIKKNAYIDELANQSPYKVTGVIENKNTYNTIKSFQD 177
QY 130 KFQDDPGLHQLDGPPLTAEDIHVHKIAARIYEENDRAVDFKIVSKLNLGLIT-ESQAH 188
Db 178 IYEGDIDTFYN-ELSSIVKEDPIDIEDTKLENLRSKIDNVYDKIQKMEIETVKSHLN 236
QY 189 LE-----DEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPEKVT 231
Db 237 IETNNKLPVILIKKIYIYDEISKELNKLEDPFNKEKLSNKISDY-DKKRQQLSEYKS 295

```


A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96796
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1871 <STO>
A:Cross-references: GB:AE005173; NID:G6143906; PIDN:AAF04452.1; GSPDB:GN00141
C:Genetics:
A:Gene: F28016.15
A:Map position: 1

```
Query Match          7.4%; Score 177; DB 2; Length 1871;
Best Local Similarity 20.7%; Pred. No. 0.71;
Matches 116; Conservative 85; Mismatches 212; Indels 148; Gaps 22;

QY 12 VLVLPQAFKP--GGSODKSLHNRSLAERPLNEQIAEAEEDKIKTYPPENKPGOSNY 69
Db 1052 VLVVEETYPKDTGGDDHNDHKEEQENVIKAEALNTEEDSFKKVEIEKQ-----1105

QY 70 SFVDNLNLLKAITKEKIEKERQIRSGPLDNKLNVEDVDSTKN---RKLDDYDSTKSG 126
Db 1106 ---DHGELKRSVMQAKRQETEEKD-KTRAMEKNETVERRKQTKDGSGLKREGEDPELGG 1161

QY 127 LDHKFQDDPGLHQLDGTPLT--AEDIVHKIAARIYEENDRAVED----KIVSKLLNLGL 180
Db 1162 --HERRGEEDRIEELVETIISDHKEKVKKKDDEYILRSQDTGKVDLGERERRSKQRKIHK 1219

QY 181 ITESQAHTEDEVAEVLQKLSKEAN-----NYEEDPNKPTSWTENQAGKIPEKVTP- 232
Db 1220 SVEDEIGDQDEDAEAAAVSVRNENGSSRKVQTIEESEK-----HKEQNKIPETSNPE 1274

QY 233 -----MAAIQDGLAKGEN-----DETVSNTLTLTNGLERRTKT 265
Db 1275 VNEEDEERVVEKETKEVEAHVQBLEGKTENCDDGGRREERGKCGMTAENMLQRQFKT 1334

QY 266 YSEDNPELQYFPNFPYALLKSID--SEKAKEKETLITIMKTLDIVKMMVKYGTI----319
Db 1335 KSDD-----GIVRKIQETKEEPDEKKS-----QESSSHVVVKLVAEEDGSLRNL 1378

QY 320 ---SPEEGVSYLENLDEMIALQTKNLEKNATD--NISKLFPAPSEKSHETDST-----369
Db 1379 EFSEKSTVSKMLKLDSEKSEKHKIRKPTERSNAPVIEKOGNKKNAEENMQDKIDRR 1438

QY 370 -KEEAAKMEKEYGSLKD-----STKDDNSNPGGKTD--399
Db 1439 GKQEIKGQBPYGLVRNGEHDKITEVHRGEEKGTAENVSTKIQTQKDELEKPRKPEIS 1498

QY 400 -----EPKGTAEYLAIRKNIEWLKXHDKKGKEDYDLS--KMRDFIN 441
Db 1499 ENNHIEFMDSSQSDIEEKGSQAEKYAKQNKIQVMNDEK--KEEYHISERVNEMA 1556

QY 442 K---QADAYVEKGILDKEEAE 459
Db 1557 KRILQVESKANDGSSKKNETE 1577
```

Search completed: January 2, 2004, 11:02:28
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 17 Seconds
(without alignments)
1294.618 Million cell updates/sec

Title: ABU66614

Perfect score: 2399

Sequence: 1 MGFLGTGTLVLVLPQAF.....EKGILDKERAEAKRYSSL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2395	99.8	468	1 SG3 HUMAN	Q8wxd2 homo sapien
2	2102.5	87.6	471	1 SG3 MOUSE	P47867 mus musculus
3	2096.5	87.4	471	1 SG3 RAT	P47868 rattus norv
4	191	8.0	2464	1 MAPB MOUSE	P14873 mus musculus
5	186.5	7.8	2116	1 MYS2 DICDI	P08799 dictyosteli
6	183.5	7.6	1790	1 US01 YEAST	P25386 saccharomyc
7	180.5	7.5	2459	1 MAPB RAT	P15205 rattus norv
8	179	7.5	1005	1 RA50 METJA	Q58718 methanococc
9	178.5	7.4	1427	1 REST HUMAN	P30622 homo sapien
10	178	7.4	882	1 RA50 PYRFU	P58301 pyrococcus
11	172.5	7.2	1164	1 BAG STRAG	P27951 streptococc
12	169	7.0	700	1 TRDN CANPA	P82179 canis famil
13	167.5	7.0	957	1 KP5C HUMAN	O60282 homo sapien
14	166.5	6.9	1875	1 MFP1 YEAST	Q02455 saccharomyc
15	166	6.9	2663	1 CENO HUMAN	Q02224 homo sapien
16	166	6.9	5038	1 PCLO MOUSE	Q9gyx7 mus musculus
17	163.5	6.8	956	1 KP5C MOUSE	P28738 mus musculus
18	163	6.8	1251	1 RBP2 PLAVB	Q00799 plasmodium
19	161.5	6.7	1360	1 CING XENLA	Q9ptd7 xenopus lae
20	161.5	6.7	1969	1 MYS4 CAEEL	P12844 caenorhabdi
21	161	6.7	852	1 RA50 THENA	Q9x1x1 thermotoga
22	159.5	6.6	997	1 SCP1 RAT	Q03410 rattus norv
23	159.5	6.6	1978	1 MYHB CHICK	P10587 gallus gall
24	159	6.6	1208	1 PCP1 SCHPO	Q92351 schizosacch
25	158.5	6.6	1526	1 MYS2 SCHPO	Q9usi6 schizosacch
26	158	6.6	1972	1 MYHB RABIT	P35748 oryctolagus
27	157.5	6.6	1972	1 MYHB HUMAN	P35749 homo sapien
28	157.5	6.6	2230	1 GOG4 HUMAN	Q13439 homo sapien
29	157	6.5	727	1 MFP1 ARATH	Q91w85 arabidopsis
30	157	6.5	944	1 NUF1 YEAST	P32380 saccharomyc
31	157	6.5	2017	1 MYSN DROME	Q99323 drosophila
32	156.5	6.5	1087	1 AKA9 RABIT	Q28628 oryctolagus
33	156.5	6.5	2468	1 MAPB_HUMAN	P46821 homo sapien

RESULT 1
SG3_HUMAN STANDARD; PRT; 468 AA.
AC Q8WXD2; Q96C83; Q96GE8; Q9Y6G7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Secretogranin III precursor (SgIII).
GN SGC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pituitary;
RA Song H., Peng Y., Huang Q., Dai M., Mao Y., Zhang Q., Mao M., Fu G.,
Luo M., Chen J., Hu R.;
RT "Human secretogranin III mRNA, complete cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=22093564; PubMed=12098761;
RT Rong Y.P., Liu F., Zeng L.C., Ma W.J., Wei D.Z., Han Z.G.;
RC "Cloning and characterization of a novel human secretory protein:
secretogranin III.";
RL Acta Biochim. Biophys. Sin. 34:411-417(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.B., Scheetz T.E.,
Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
secretory granules.
CC -!- TISSUE SPECIFICITY: Expressed in brain, heart, kidney, liver and
skeletal muscle.

34 156 6.5 705 1 TRDN_RABIT
35 155.5 6.5 1130 1 YL17_CAEEL
36 155.5 6.5 5085 1 PCLO RAT
37 154.5 6.4 1539 1 Y373 HUMAN
38 154 6.4 1433 1 REST_CHICK
39 154 6.4 1957 1 SPOF_SCHPO
40 154 6.4 1962 1 MYS4_DROME
41 153.5 6.4 1976 1 MYHA RAT
42 153.5 6.4 2104 1 MYS3_SCHPO
43 152.5 6.4 2022 1 ANTI_ONCVO
44 152.5 6.4 8545 1 ANCI_CAEEL
45 152 6.3 667 1 CYLI_BOVIN

ALIGNMENTS

Q28820 oryctolagus
Q11102 caenorhabdi
Q9jks6 rattus norv
O15078 homo sapien
O42184 gallus gall
Q10411 schizosacch
P05661 drosophila
Q9j102 rattus norv
Q14157 schizosacch
P21249 onchocerca
Q9n4m4 caenorhabdi
P35662 bos taurus

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95001263; PubMed=7917832;
RA Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA Sutcliffe J.G.;
RT Primary structure of mouse secretogranin III and its absence from
RT deficient mice.;
RL J. Mol. Neurosci. 4:225-233(1993).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=90376160; PubMed=2204698;
RA Ottiger H.-P., Battenberg E.F., Teou A.-P., Bloom F.E.,
RA Sutcliffe J.G.;
RT "B1075: a brain- and pituitary-specific mRNA that encodes a novel
RT chromogranin / Secretogranin-like component of intracellular
RT vesicles.";
RL J. Neurosci. 10:3135-3147(1990).
CC -I- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC secretory granules.
CC -I- TISSUE SPECIFICITY: BRAIN- AND PITUITARY-SPECIFIC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U02983; AAA56637.1; --
KW Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 471 SECRETOGRANIN III.
SQ SEQUENCE 471 AA; 53183 MW; 977B3F885F33223 CRC64;
Query Match 87.4%; Score 2096.5; DB 1; Length 471;
Best Local Similarity 87.9%; Pred No. 1.1e-93;
Matches 414; Conservative 18; Mismatches 36; Indels 3; Gaps 1;
QY 1 MGFLGTGWLVLVL---PIQAFPPGSGQSKLHNRLSALRPLNEQIAEAEADKIKKT 57
Db 1 MGFLWTGSWLVLVLNSGPIQAFPPGSGQSKLHNRLSALRPLNEQIAEAEADKIKKT 60
QY 58 YPPENKPGQSNYSFVDNLNLKAITKEKIKEROSIRSSPLDNKLVNVEDVDSTNKRKLI 117
Db 61 YPSEKPSSESNFSSVDNLNLKAITKETVETKAKQSISSPFDNRLNVDADSTNKRKLT 120
QY 118 DDYDSTKSLDGHKFDODPDGLHOLDGTPLTAEIVHKAARIYENDRAVFDKIVSKLLN 177
Db 121 DEYDSTKSLDGHKFDODPDGLHOLDGTPLTAEIVHKAARIYENDRAVFDKIVSKLLN 180
QY 178 LGLITESQAHTLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIPKVPMAAIQ 237
Db 181 LGLITESQAHTLEDEVAEVLQKLSKEANNYEAEPEKTSRTENQDGKIPKVPVAAATQ 240
QY 238 DGLAKGENDETVSNLTITNGLERTKTSYSDNFEELYFPNFPVALLKSIDSEKEAKE 297
Db 241 DGFNTRENDTTSVNTLTLSNGLERTNPHRDDFEELYFPNFPVALLTIDSEKEAKEKE 300
QY 298 TLITIMKTLIDPVKMWKYGITISPEEGVSYLENLDEMIALQTKNLEKNADNTSKLPFA 357
Db 301 TLITIMKTLIDPVKMWKYGITISPEEGVSYLENLDEMIALQTKNLEKNADNTSKSKLPFA 360
QY 358 PSEKSHETDSTKEAAAKWEKYSGLKSDTKDDNSNPGCKTDEPKGKTEAYLEAIRKNIE 417
Db 361 PPEKSHETDSTKEAAAKWEKYSGLKSDTKDDNSNPGCKTDEPKGKTEAYLEAIRKNIE 420
QY 418 WLKXGDKKGNKEDYDLSKQKRDPIFNQADAYVEKGILDKKEAEAIKRIYSSL 468

Db 421 WLKXHNKGNKEDYDLSKQKRDPIFNQADAYVEKGILDKKEAEAIKRIYSSL 471
RESULT 4
MAPB MOUSE
ID MAPB MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAPI(X))
DE [Contains: MAPI light chain LC1].
GN MAPIB OR MTAPIB OR MTAP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DOMAIN.
RC STRAIN=Swiss Webster; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "The microtubule binding domain of microtubule-associated protein
RT MAPIB contains a repeated sequence motif unrelated to that of MAP2
RT and tau.";
RL J. Cell Biol. 109:3367-3376(1989).
CC -I- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAPIB MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAPIB BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -I- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAPIB AND MAPIB PROTEINS.
CC -I- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V; repeated but not at fixed intervals, which is
CC responsible for the binding of MAPIB to microtubules.
CC -I- PTM: LC1 IS COEXPRESSED WITH MAPIB. IT IS A POLYPEPTIDE GENERATED
CC FROM MAPIB BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAPIB AND MAPIB. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAPIB.
CC -I- SIMILARITY: TO MAPIA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X51396; CAA35761.1; --
DR PIR; S07549; QRMSP1.
DR MGD; MGI:1306778; Mtap1b.
DR GO; GO:0016358; P:dendrite morphogenesis; IMP.
DR GO; GO:0001578; P:microtubule bundling; IMP.
DR InterPro; IPR000102; MAPIB neuraxin.
DR Pfam; PF00414; MAPIB neuraxin; 10.
DR PROSITE; PS00230; MAPIB_NURAXIN; 7.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAPI LIGHT CHAIN LC1.
FT REPEAT 1874 1890 MAPIB 1.
FT REPEAT 1891 1907 MAPIB 2.
FT REPEAT 1908 1924 MAPIB 3.
FT REPEAT 1925 1941 MAPIB 4.
FT REPEAT 1942 1958 MAPIB 5.
FT REPEAT 1959 1975 MAPIB 6.
FT REPEAT 1993 2009 MAPIB 7.
FT REPEAT 2010 2026 MAPIB 8.
FT REPEAT 2027 2043 MAPIB 9.
FT REPEAT 2044 2060 MAPIB 10.
FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	DR	EMBL; M14628; AAA33227.1; -	FT	TURN	143	145
CC	DR	PIR; A26655; A26655.	FT	HELIX	148	150
CC	DR	PDB; 1MMA; 03-DEC-97.	FT	HELIX	155	169
CC	DR	PDB; 1MMD; 17-AUG-96.	FT	TURN	170	170
CC	DR	PDB; 1MWG; 03-DEC-97.	FT	STRAND	173	179
CC	DR	PDB; 1MWN; 03-DEC-97.	FT	TURN	181	182
CC	DR	PDB; 1MND; 17-AUG-96.	FT	TURN	185	200
CC	DR	PDB; 1MNE; 17-AUG-96.	FT	HELIX	210	226
CC	DR	PDB; 1VOM; 23-DEC-96.	FT	STRAND	227	228
CC	DR	PDB; 1LVK; 28-JAN-98.	FT	TURN	234	235
CC	DR	PDB; 1DOX; 20-DEC-00.	FT	STRAND	236	237
CC	DR	PDB; 1DOY; 20-DEC-00.	FT	STRAND	240	247
CC	DR	PDB; 1DIA; 20-DEC-00.	FT	TURN	249	250
CC	DR	PDB; 1DIB; 20-DEC-00.	FT	STRAND	253	261
CC	DR	PDB; 1D1C; 20-DEC-00.	FT	HELIX	265	268
CC	DR	PDB; 1FMV; 20-DEC-00.	FT	TURN	269	269
CC	DR	PDB; 1FMW; 20-DEC-00.	FT	TURN	273	274
CC	DR	PDB; 1G8X; 17-JAN-01.	FT	STRAND	278	278
CC	DR	PDB; 1JWY; 07-NOV-01.	FT	HELIX	279	287
CC	DR	PDB; 1JX2; 07-NOV-01.	FT	TURN	290	296
CC	DR	DictyDb; DD01008; mhca.	FT	TURN	297	297
CC	DR	InterPro; IPR000048; IQ region.	FT	STRAND	301	303
CC	DR	InterPro; IPR001609; myosin head.	FT	TURN	305	307
CC	DR	PIfam; PF00612; IQ; 2.	FT	TURN	316	317
CC	DR	PIfam; PF02736; Myosin N; 1.	FT	HELIX	320	334
CC	DR	PRINTS; PR00193; MYOSINHEAV.	FT	TURN	335	335
CC	DR	ProDom; PD000355; myosin head; 1.	FT	TURN	356	356
CC	DR	SMART; SM00015; IQ; 1.	FT	STRAND	360	360
CC	DR	SMART; SM00242; MYSC; 1.	FT	STRAND	368	368
CC	DR	PROSITE; PS0096; IQ; 1.	FT	HELIX	373	382
KW	MYOSIN; Coiled coil; Actin-binding; ATP-binding; 3D-structure;					
KW	Calmodulin-binding; Methylation; Alkylation; Phosphorylation.					
FT	DOMAIN	1 761	QY	26	SDQS--LHNRELSAERPLNEQIAEEDKIKTTPPENPGQNSYFVDMNLLKKAITE	83
FT	DOMAIN	762 791	DB	1023	TKDSSELLRQKKLEELKQVEALAAETAAKLAQAANKKLOQGYTEL-NEKFNSEVTA	1081
FT	DOMAIN	791 817	QY	84	KEKTEKERQSTRSS--PLDNKLNVEDVDSTNKRKLIDDYDSTKSLGDKHKKFODDPDGLHQL	141
FT	DOMAIN	817 2116	DB	1082	RSNVESKSKTLESOLVAVNNELDEE---KKNR---DALEKKKALDAMLEEMKDKQLEST	1134
FT	NP BIND	179 186	QY	142	DGTPLTAEIVHKIAARIYEENDRAVDFKIVSKLNLGLITESQAHTLEDEVAEVLQKLI	201
FT	DOMAIN	638 660	DB	1135	GGEKSLYDL-----KVQESDMEALRNQISELQSTIAKLEKIKSTLEGEVARLQGELE	1188
FT	MOD RES	130 130	QY	202	SKE--ANNYEDPNKPTSWTENQAGKIPEKVT-----PMAAIQDGLAKGEN	245
FT	MOD RES	1823 1823	DB	1189	AEQLAKSNVEKQKKKVELDEKSAQLAEBETAQAALDKLKKLEQBLSEVQTQLSEANN	1248
FT	MOD RES	2029 2029	QY	246	D-----ETVSNLTILTNGLERRTKTYSEDN----PEELQYFPNFVALLKSIDSE	290
FT	TURN	3 5	DB	1249	KVNSDSTNKHLETSPNNKLELEAEQAKQAOKALEKKLGLSEELKH-----VNEQLEEE	1302
FT	TURN	7 8	QY	291	KEAKEK-----ETLITIMKTLDVQVMVKYGTISPSEGVSYLE-----	329
FT	HELIX	10 15	DB	1303	KKQESNEKRVLEKEVSELKQDIEEVASKKAVTEAKKKKSESELDEIKRQYADVSSR	1362
FT	STRAND	34 37	QY	330	--NLDEMIALQTN-KLEKNATONISKLPAPSPKSHEETD-----STKEEAAKMEKEY	380
FT	TURN	41 42	DB	1363	DKSVEQLTKLQAKNVEELRNATAEEAEGQLDRAERSKKKAEFDLEAVKNLEEBEATAKKVKA	1422
FT	STRAND	59 63	QY	381	GSLLKSTKDDNSPGGKTDEPKGTEYLAIRKNIWLVKHKDKKGNKEDYDLKMDFI	440
FT	STRAND	69 73	DB	1423	KAMKKAETDYRSTKSELDKANSVSSEQYVQIKRLNEE-----LSELRSVL	1467
FT	STRAND	74 76	QY	441	NKQADAVVEKGILDKEAEA	460
FT	STRAND	78 79	DB	1468	-EADERCNSAIKAKKTAES	1486
FT	STRAND	83 85				
FT	STRAND	86 87				
FT	STRAND	90 90				
FT	STRAND	91 93				
FT	STRAND	99 110				
FT	STRAND	111 113				
FT	STRAND	116 119				
FT	STRAND	120 121				
FT	STRAND	122 126				
FT	STRAND	137 142				

RESULT 6
 US01_YEAST STANDARD; PRT; 1790 AA.
 AC P25366;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein US01.
 GN US01 OR INT1 OR YDL058M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RX MEDLINE=91185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yanasaki M.;
 RT "A cytoskeleton-related gene, us01, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 113:245-260(1991).
 RN [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X54378; CAA38253.1; -;
 CC EMBL; L03188; AAB00143.1; -;
 CC EMBL; U53668; AAB66659.1; -;
 CC EMBL; U53668; AAB66659.1; -;
 CC SGD; S0002216; US01.
 CC InterPro; IPR002017; Spectrin.
 CC InterPro; IPR006955; Usol_p115_C.
 CC InterPro; IPR006953; Usol_p115_head.
 CC Pfam; PF04871; Usol_p115_C_1.
 CC Pfam; PF04869; Usol_p115_head; 1.
 CC Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 KW DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 1 724
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 I -> V (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).

SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
 Query Match 7.6%; Score 183.5; DB 1; Length 1790;
 Best Local Similarity 20.3%; Pred. No. 0.15;
 Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;
 QY 29 KSLHN--RELGAERPLNQIAEAEEDKIKTKYPPENKPGQSNYSFVDNLLKAI---T 82
 DB 965 KSLANNYKMQAE--NESLIKAVEE-----SKNESSIQLSNLKNKIDMSQ 1008
 QY 83 EKEIEKERQIRSSPLDNKLVNVDSTNKRKLI-----DDYSTKSGLDHFKFD--- 133
 DB 1009 EKENFQIERGSIKIEKIQKKTISDLQTEKBEIISKSDSKDEYESQISLLKEKLETTAT 1068
 QY 134 -DPDLGLHLDGTPLTAEDIVHKIARIYEENDRAVDFDKIVSKLLNLGLITESQAH----- 187
 DB 1069 ANDENVNKISBELTKTREETLEAEALAA--YKNLKNLELTLETSEKALKVEKNEEHLKBEK 1126
 QY 188 -TLEDEVAEVLQKLISKAN-----NYEED--PNKPTSWTNOAGKIPEK 229
 DB 1127 IQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKEQYNE--EISQND 1185
 QY 230 VTPMAAIQDGLAKGENDETVNTLTNGLERRT-----KTYSEDN-- 270
 DB 1186 ITSTQQENESIKK-KNDELEGEVRAMKSTSEQSNLKKSEIDALNLQIKELKKQNETNEA 1244
 QY 271 -----FELQYFPNP-----YALLKSIDSEKEAK 294
 DB 1245 SLLESIKVSESETVKIKELQDECNPKFKEVSELEDKLKASEDKNSKYLELQK-ESEKIKE 1303
 QY 295 EKETLIITIMKTIDFVKMMVK-----SESELRLKTSSEERKNAEQLKNEIQIKNQ 1363
 DB 1304 ELDAKTTTELKIQLEKINTLSKAKSESELSRLKTSSEERKNAEQLKNEIQIKNQ 1363
 QY 316 -----YGTISPE--EGVSYLENDEMIALQTFNKLKKNATNISKLPAPSEK 361
 DB 1364 AFEKERKLLNCGSSITTOEYSEKINTLE--DELIRLQNELKAKEIDNTRSELEKVSLS 1421
 QY 362 SHEETDSTKEAAKMEKEVGSGLKDS-TKDDNSNPGKTDPE-----KGTAYLEAIRK 414
 DB 1422 NDELEEKQNTIKLQDLISYKOKITNDKLSIERDNKRDLSESLKQLRAAESKAK 1481
 QY 415 NIEWLKHKDKGNKEDVDLSKVRDFINKQADAYVEKGILD-KEEAIAIKR 463
 DB 1482 VEEGLKXLEBESSKEAELEKSKEMW-KKLESTIESNETELKSSMETIRK 1530
 RESULT 7
 MAPB RAT STANDARD; PRT; 2459 AA.
 AC P15205; O62958; Q9ER21; Q9QW92;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
 DE light chain LC1].
 GN MAP1B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-142 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;
 RX MEDLINE=96257242; PubMed=8666295;
 RA Liu D., Fischer I.;
 RT "Isolation and sequencing of the 5' end of the rat microtubule-
 RL associated protein (MAP1B)-encoding cDNA.";
 RL Gene 172:307-308(1996).
 RN [2]
 RP SEQUENCE OF 96-2459 FROM N.A.; DOMAIN, AND INDUCTION.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
 RX MEDLINE=92347374; PubMed=1639092;

RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
 RT "Identification of two distinct microtubule binding domains on
 RL recombinant rat MAP 1B.";
 RN Eur. J. Cell Biol. 57:66-74 (1992).
 [3]

RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.

RC TISSUE-Spinal cord;

RX MEDLINE=90059871; PubMed=255150;

RA Rientz A., Greening G., Hermans-Borgmeyer I., Kirsch J.,
 RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
 RT "Neuraxin, a novel putative structural protein of the rat central
 RT nervous system that is immunologically related to microtubule-
 RT associated protein 5.";
 RL EMBL J. 8:2873-2888 (1989).

RN [4]

RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.

RX MEDLINE=97405699; PubMed=9260743;

RA Ma D., Nothias F., Boyne L.J., Fischer I.;

RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
 RT in rat CNS and PNS during development.";
 RL J. Neurosci. Res. 49:319-332 (1997).

CC -!- FUNCTION: The function of brain MAPS is essentially unknown.
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
 CC that accompany neurite extension. Possibly MAP1B binds to at least
 CC two tubulin subunits in the polymer, and this bridging of subunits
 CC might be involved in nucleating microtubule polymerization and in
 CC stabilizing microtubules.

CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
 CC with MAP1A and MAP1B proteins.
 CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
 CC heart or muscle.

CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
 CC nerve levels are high early in development but decrease during
 CC postnatal development and are low in adults. In dorsal root
 CC ganglia levels remain high throughout development.
 CC -!- INDUCTION: By nerve growth factor.

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.

CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (by similarity).

CC -!- PTM: Phosphorylated.

CC -!- SIMILARITY: TO MAP1A.

CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to

CC 2459) was originally described as neuraxin in Ref.3.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U52950; AAB17068.1; -

DR EMBL; X60370; CAC16162.1; -

DR EMBL; X16623; CAA34620.1; ALT_SEQ.

DR PIR; A56577; A56577.

DR Pfam; IPR000102; MAP1B neuraxin.

DR Pfam; PF00414; MAP1B neuraxin; 10.

DR PROSITE; PS00230; MAP1B NEURAXIN; 8.

KW Microtubules; Repeat; Phosphorylation.

FT CHAIN ? 2459 MAP1 LIGHT CHAIN LC1.

FT REPEAT 1869 1885 MAP1B 1.

FT REPEAT 1886 1902 MAP1B 2.

FT REPEAT 1903 1919 MAP1B 3.

FT REPEAT 1920 1936 MAP1B 4.

FT REPEAT 1937 1953 MAP1B 5.

FT REPEAT 1954 1970 MAP1B 6.

FT REPEAT 1988 2004 MAP1B 7.

FT	REPEAT	2005	2021	MAP1B 8.
FT	REPEAT	2022	2038	MAP1B 9.
FT	REPEAT	2039	2055	MAP1B 10.
FT	DOMAIN	559	1035	GLU-RICH
FT	DOMAIN	588	786	LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).
FT	DOMAIN	2224	2312	LYS-RICH
FT	CONFLICT	127	127	M -> V (IN REF. 1).
FT	CONFLICT	140	140	T -> S (IN REF. 1).
FT	CONFLICT	2112	2112	R -> K (IN REF. 3).
FT	CONFLICT	2169	2169	L -> I (IN REF. 3).
SQ	SEQUENCE	2459 AA;	269497 MW;	2E3F6872DEDB8BA2 CRC64;

Query Match 7.5%; Score 180.5; DB 1; Length 2459;
 Best Local Similarity 22.7%; Pred. No. 0.31;
 Matches 120; Conservative 77; Mismatches 200; Indels 131; Gaps 22;

QY 22 KPGSQDQSLHNLRLSALRPLNEICIAEEDKIK-KTYPPEKPGQSNYSVDNLNLLKA 80
 Db 549 KPLSSKVRKSEKEEAPKASQVEKTPKVESKEKIVVKDKPKG-----VESKPS 600
 QY 81 ITEKEIEKERQSRSSPLONKLNVEDVDSTNKRKLDDYDSTKSLDGHKFDQDDP--DGL 138
 Db 601 VTEKEVSEKEEQSPVKAEEAKAATESKPKVKTKDKVVKKEIKTKP--EEKKEEKPKKEVA 658
 QY 139 HQLDGTPLTADIVHKIAARIYENDRAVFDKIVSKLNLNLGLITESQAHT-LEDEVAEVL 197
 Db 659 KKEDKTPLKDKDEPKK-----EAKKEIKKEIKKEEKELKKEKVKETPLDKAKKEV- 710
 QY 198 QKLISKANNEYEDPNK-----PTSTWENQAG-----KIPEKVTPTMAA- 235
 Db 711 KQDEKKEVKEEKEPKKEIKKIKKIDKISDTKPLSDTKKPAALKPKVAKKEEPTKKEPTAA 770
 QY 236 -IQD-GLAK-----GENDETVSNTL-----TLTNGLERRTK 264
 Db 771 KLKDKGKVKVKEGKTEAAATAGTAAVAAAGVAAAGPAKLEAERSLMSSEPDLT 830
 QY 265 TYSEDNFELQYFPNLYALLKSIDSEKAKE-----KETLIITIMKTLIDFVKMMV 314
 Db 831 DFELKAEEDIVAKDIKPOLLEIEDEKLKETEPEGAUVIQKETEVS-----877
 QY 315 KYGTISPERGVSYLENLDEMIALQTKNL---EKVATDNISKL-----PPAPSEKSHEE 365
 Db 878 KGSAESDEGITTTTEGE--CEQTPPELFPVEKQGVDDIEKFEDEGAESESEAGDYE 935
 QY 366 TDSKKEAAKWEKEYGSLKSDTKDDNSNPGKTDPEKGTAE--YLEAIRKNI-----416
 Db 936 EKAETEAEPEEDG-----EDNVSGSASKHSPTDEETAKAEADVHIKKEKESVAGSDRA 992
 QY 417 -----EWLKHKKKKNKEDYDLKSMRDFINKQADAVYVEKGLDKKEAE 459
 Db 993 EEDMDALEKGEAFQSEEGEEES-----DKAEDAREDEHPDKTEAE 1035

RESULT 8

RA50 METJA

ID RA50 METJA STANDARD; PRT; 1005 AA.

AC Q58718;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA double-strand break repair rad50 ATPase.

GN RAD50 OR MJ1322.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococaceae; Methanocaldococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.,
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
 CC rad50/mre11 complex possesses single-strand endonuclease activity
 CC and ATP-dependent double-strand-specific exonuclease activity.
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding
 CC and/or repositioning DNA ends into the mre11 active site (By
 CC similarity).
 CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U67572; AAB99331.1; -.
 DR TIGR; MJ1322; -.
 DR HAMAP; MF_00449; -; 1.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR003405; SMC_C.
 DR InterPro; IPR003395; SMC_N.
 DR Pfam; PF04423; Rad50_zn_hook; 1.
 DR Pfam; PF02483; SMC_C; 1.
 DR Pfam; PF02463; SMC_N; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
 FT NP BIND 32 39 ATP (BY SIMILARITY).
 FT DOMAIN 158 849 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E78F3 CRC64;
 Query Match 7.5%; Score 179; DB 1; Length 1005;
 Best Local Similarity 22.3%; Pred. No. 0.12;
 Matches 116; Conservative 80; Mismatches 164; Indels 160; Gaps 27;
 QY 52 DKIKKTYPPENKQSNYSFVDNLLKA-ITEKEIEKEKQSTRSSPLDNKLN-VEDVD 109
 Db 159 DEFEKCY---QKMGIEVKEKRLERIEGELNYKENYEKE-----LKNKMSQLEE-- 205
 QY 110 STNRKLIDVDSTKSGLDHFKQDDPGLHQLDGTPLTAEDIVHKAARIYEENDRAVFD 169
 Db 206 --KNKLMELNDKLNK-IKEFEFIEKLFENWENKLLYEKFNK-----EEKRALEL 257
 QY 170 KIVS-KLNLNGLITESQAH-----TLEDEVAEVLQKLSKEANNYEEDPNK 214
 Db 258 KNQELKILEYDLNVTVEARETLNRHKDEYEKYSLSVDIRKIESRL--RELKSHVEDYLK 315
 QY 215 PTSWENQAKIPKVTPTMAIQDLAKGENDETVNTILTNGLERTKTYTSEDNFEEL 274
 Db 316 LTKQLEIKIGDI-EK-----LKEFINKSKYRDDIDNLDLTILNKI-----KDEIERV 360
 QY 275 QYFPNFVALLKSIDSEKEAKEK-----ETLITITMKTILDFKVMVKYGTISPE 322
 Db 361 ETIKLLEELNLENEEIEKIEKIKRICEBEKYEKYLEEFAVENKUTLEYITL--- 417
 QY 323 EGVSYLENLDEMIALQPNKLEKNATD---NISKLPFAPS-----EKSHETDSTK-- 370
 Db 418 -----LQEKKSIEKINDLETRINKLLETRINKLLETRINKLLETRINKLLETRINKL 463
 QY 371 -----EBAKMEKEYGSLKSTK-----DDNSNPGK-----TDEPK-----GKT 405

Db 464 LENLQKEIEIENKGLGEINSEIKRLKILDELKEVGRKPLCKTPTIDENKMKELINQHT 523
 QY 406 E---AYLE-----AIRKNIEWLKHH-----DKGKNKEDYDI----- 433
 Db 534 QLNKKYTELEIEINKIRIEIKDEIKLKEIKDEENLKTLYLEKQSIIELELKLKNY 583
 QY 434 SKMRDFINKQADAVVEKG-----ILD--KEEAIAIKRYS 466
 Db 584 KEQLDEINKKISNYVINGKVPVDEILEDIKSQLNKFKNFYN 623
 RESULT 9
 ID REST_HUMAN STANDARD; PRT; 1427 AA.
 AC P30622;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
 DE Sternberg intermediate filament associated protein).
 GN RSN
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood monocytes;
 RX MEDLINE=92289675; PubMed=1600942;
 RA Bilbe G., Delabie J., Brueggen J., Richener H., Aselbergs F.A.M.,
 RA Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
 RA de Wolf-Peters C., Shipman R.;
 RT "Restin: a novel intermediate filament-associated protein highly
 RT expressed in the Reed-Sternberg cells of Hodgkin's disease.";
 RL EMO J. 11:2103-2113(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92405160; PubMed=1356075;
 RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
 RA "CLIP-170 links endocytic vesicles to microtubules.";
 RL Cell 70:887-900(1992).
 CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
 CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
 CC CYTOSKELETON
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P30622-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P30622-2; Sequence=VSP_000765;
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
 CC OF HODGKIN'S DISEASE.
 CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X64838; CAA46050.1; -.
 DR EMBL; M97501; AAA35693.1; -.
 DR PIR; S22695; S22695.
 DR Genew; HGNC:10461; RSN.
 DR MIM; 179838; -.
 DR GO; GO:0005768; C:endosome; TAS.
 DR GO; GO:0005882; C:intermediate filament; TAS.
 DR GO; GO:0015630; C:microtubule cytoskeleton; TAS.
 DR GO; GO:0008017; F:microtubule binding activity; TAS.

[illegible]

```

[1] SEQUENCE FROM N.A., AND CHARACTERIZATION.
    STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
    MEDLINE=20485553; PubMed=11029422;
    Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
    Carney J.P.;
    "Mre11 and Rad50 from Pyrococcus furiosus: cloning and biochemical
    characterization reveal an evolutionarily conserved multiprotein
    machine.";
    J. Bacteriol. 182:6036-6041(2000).
[2] SEQUENCE FROM N.A.
    STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
    Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
    "The complete sequence of the Pyrococcus furiosus genome.";
    Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
[3] X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
    STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
    MEDLINE=20348838; PubMed=10892749;
    Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
    Carney J.P., Tainer J.A.;
    "Structural biology of Rad50 ATPase: ATP-driven conformational
    control in DNA double-strand break repair and the ABC-ATPase
    superfamily.";
    Cell 101:789-800(2000).
    !- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
    rad50/mre11 complex possesses single-strand endonuclease activity
    and ATP-dependent double-strand-specific exonuclease activity.
    Rad50 provides an ATP-dependent control of mre11 by unwinding
    and/or repositioning DNA ends into the mre11 active site.
    !- SUBUNIT: Forms a complex with mre11.
    !- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
    -----
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    -----
    EMBL; AE010225; RAL81291.1; -.
    PDB; 1F2T; 20-SEP-00.
    PDB; 1F2U; 02-AUG-00.
    PDB; 1I18; 30-MAY-01.
    PDB; 1I8D; 28-AUG-02.
    HAMAP; MF_00449; -. 1.
    InterPro; IPR003439; ABC transporter.
    InterPro; IPR003395; SMC_N.
    Pfam; PF04423; Rad50_zn_hook; 1.
    Pfam; PF02463; SMC_N7; 1.
    ProDom; PD000006; ABC transporter; 1.
    DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
    Complete proteome.
    NP_BIND 30 37 ATP.
    DOMAIN 148 744 COILED COIL (POTENTIAL).
    SEQUENCE 882 AA; 103839 MW; 3ADCBD250382A99E CRC64;
    Query Match 7.4%; Score 178; DB 1; Length 882;
    Best Local Similarity 22.1%; Pred. No. 0.12;
    Matches 112; Conservative 90; Mismatches 172; Indels 132; Gaps 21;
    QY 39 ERLPNEQTAEE--DKYK---KTYPPENKPGQSIVSFVDNLNLKA-----ITEKEKIE 88
    DB 260 KTLREIRKNTBEYLKLEKEKELEEQVKETISIKQVDAYLAKFEKNEVLDKKYKIE 319
    QY 89 KERQSIRSSPLDKNLNVEDVDSKTKRKLDDYDSTKSGLDHFKQDDPDGLHQLDGLPLTA 148
    DB 320 KE--ITRVEELNIEIQ-----KRIELNEKESEKEKLENE-----K 353
    QY 149 EDIVHKIAARIYEENDRAVDFDKIVSKLNLGLITESQAHTLDEVAEVLQKLISKANNY 208

```

```

Db 354 KEILNKLAIL---EKDQHYBEIKAKKENLQKKEKLGKSPEDIKKLEBELETKK-TTI 409
QY 209 EEDPNKPTSWTENOAGKIPEKVTWMAAIQDGL--AKGS-----NDE----- 247
Db 410 EERNEIT-----QRIGELKNIGDKLTAIEELKAKGKCPVCGRELTDHREELLSKYHL 465
QY 248 -----TVSNLTLTNGLERTKTYSNDFEELQYFPNFYALLKSIDS-----E 290
Db 466 DLNNSKNTLAKLIDRKSELERLRIDME-IKRLTPLLTVAEQIRSIIEELNVNVLSEKIE 524
QY 291 KEAKEKETLITIMKTLDVFMVVKYGTISPEBGVSYLENDEMIA--LQTKNKLKQAT 348
Db 525 KNATEYKELBELTLEGRIRGLA-----EDLKKLAPLEKKLAALIHKKQLEKELK 576
QY 349 DNISKLPAPSEKSHETDSTKEBAAKMEKEYGSLKSTKD-----DNSNPGGKTD 399
Db 577 ELNTKL-ESFGFKSVEDLDSKLRELEIKYKLYTLNLSKKELEITQREIAKAKETLEMSF 635
QY 400 EPKGTTRAYLEAIRKNIEWLKXHKHKKGNKEDY-----DLSKMR 437
Db 636 BELAEVADIERIEKLSQLK---QKYNEEYKKKREEKELEKELARLEAQKKELEKRR 692
QY 438 DFINKQADAYVEKGILDKEEAIAKR 463
Db 693 DTIK-----STLEKKAEBKENRERVKK 714

RESULT 11
BAG_STRAG
ID_BAG_STRAG STANDARD; PRT; 1164 AA.
AC P27551;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE IGA FC receptor precursor (Beta antigen) (B antigen).
GN BAG.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN 11;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
RC STRAIN=LA239;
RX MEDLINE=91312121; PubMed=1857207;
RA Jerlstroem P.G., Chhatwal G.S., Timmis K.N.;
RA Bateman A., Eddy S.R., Chothia C.;
RT "The Iga-binding beta antigen of the c protein complex of Group B
RT streptococci: sequence determination of its gene and detection of two
RT binding regions."
RL Mol. Microbiol. 5:843-849(1991).
RN [2]
RP IDENTIFICATION OF IG-LIKE DOMAIN.
RX MEDLINE=97035265; PubMed=8880921;
RA Bateman A., Eddy S.R., Chothia C.;
RT "Members of the immunoglobulin superfamily in bacteria."
RL Protein Sci. 5:1939-1942(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (potential).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59771; CAA2442.1; -.
CC PIR; S15330; FCSOAG.
CC InterPro; IPR004829; Csurface antigen.
CC InterPro; IPR005877; Gpos YSIRK.
CC InterPro; IPR001899; Gram_pos_anchor.
CC InterPro; IPR003599; Ig.

```

```

InterPro; IPR006192; LPXTG.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF05062; RICH; 1.
DR Pfam; PF04650; YSIRK_signal; 1.
DR Pfam; PF053432; Csurface_antigen; 1.
DR SMART; SM00409; IG; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00835; IG_LIKE; FALSE_NEG.
KW Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
KW Immunoglobulin domain.
FT SIGNAL 1 37
FT CHAIN 38 1135 IGA FC RECEPTOR.
FT PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 434 534 IG-LIKE.
FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).
FT DOMAIN 827 945 PRO-RICH REPEATS.
FT SITE 1132 1136 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;

Query Match 7.2%; Score 172.5; DB 1; Length 1164;
Best Local Similarity 21.6%; Pred. No. 0.3;
Matches 124; Conservative 77; Mismatches 189; Indels 185; Gaps 25;

QY 36 LSAERPLNE-----QIAEAEEDKIKKTYPENKPGOSYFVDNLLMLLKAIT-EKEK 86
DB 149 LELENQFNETNRLHLHKQHEEVEKDKAK-----QOKTLKQSDTKVDLSNIDKELNHSQ 204
QY 87 IEK--EROSIRSSPLDNKL-----NVEDVDSTKNRKLIDDYDSTKSLDHLK 130
DB 205 VEKMAEQKGTINEDKXSMKLIEDIRKQAQADKEDAEVKVRELKGLFSSTKAGLDQ 264
QY 131 FQDDPDGLHQLDGTPLTADIVHKI-----AARIYEENDRAVDFKIVSKLLNLG 179
DB 265 IQE-----HVKKET--SSEENTQKVDHANSQNLAKSLLEEDKATTEGATQVKQKF 317
QY 180 L-----ITESAHTLEDEVAEV-----LQKLISKE---ANNYEE 210
DB 318 LENAQKLEIQPLIKETNVKLYKAMSESLQVEKELKENSEANLDELVAKSEIVREYEG 377
QY 211 DPNKPTSWTE-----NOAGKIPEKVTWMAAIQDGLAKGENDETVPN 251
DB 378 KLNQSKNLPKQLKEEAHSLKQVVEDEFKFKTSEQVTPKRVKRDLANENNQ---Q 434
QY 252 TLTTLNGLERTKTYSEDNFEELQYFPNFYALLKSIDSKEAKEKETLITIMKTLDVFK 311
DB 435 KIELTVSPENITVEGED-----VKFTVTAKS-DS-----KTLDFSD 471
QY 312 MMVYKG-TISPPEGVSYLENLD-----EMIALQTK-----NKLEKNATD 349
DB 472 LUTKYNPSVDRISTNYKTNTDNHKLAIITIKNLKNSQITVLKAKDDSGNVVEKFTFI 531
QY 350 NISKLPAPSEKSHETDSTKEEAA-----KMEKEYGSLKD 385
DB 532 TVQKKEEQKVPKTPQDKDSTKEEVQPEKPNQKQLELIKSAQOELEKLEKALME 591
QY 386 STKDDNSNPGKTDPEPKKTEYLAIRKNTIEWLK-----HDKGNKEDYDLSKNRDFIN 441
DB 592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYTYTHYFNKYKSDPM 648
QY 442 KQADAYVEKGILDKEEAEB-----AIKEIYSS 467
DB 649 YQLHAQME--MLTRKVVQVMNKNYPDNABEIKKIPES 681

RESULT 12
TRDN CANFA
ID TRDN CANFA STANDARD; PRT; 700 AA.
AC P82179;
DT 28-FEB-2003 (Rel. 41, Created)

```


CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE). A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB011103; BAA25457.1; -
CC EMBL; AF010146; AAD01436.1; -
CC HSSP; P56536; 2KIN.
CC Genew; HGNC:6325; KIF5C.
CC MIM; 604593; -
CC GO; GO:0005871; C:kinesin complex; TAS.
CC GO; GO:0003777; F:microtubule motor activity; TAS.
CC GO; GO:0006996; P:organelle organization and biogenesis; TAS.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Microtubules; ATP-binding; Coiled coil.
CC DOMAIN 1 386 KINESIN-MOTOR (BY SIMILARITY).
CC DOMAIN 406 923 COILED COIL.
CC DOMAIN 859 956 GLOBULAR.
CC DOMAIN 174 315 MICROTUBULE-BINDING.
CC NP_BIND 86 93 ATP (BY SIMILARITY).
CC CONFLICT 355 360 TLKVI -> STHAVI (IN REF. 2).
CC CONFLICT 583 585 EFT -> DRV (IN REF. 2).
CC SEQUENCE 957 AA; 109494 MW; A9F25BBIC994322A CRC64;
CC
CC Query Match 7.0%; Score 167.5; DB 1; Length 957;
CC Best Local Similarity 20.3%; Pred. No. 0.42;
CC Matches 103; Conservative 98; Mismatches 182; Indels 125; Gaps 23;
CC
CC 25 GSQDKSLHNR-----ELSAERPLNEQIAEAEEDKIKKTPPENKPGQSNYSFVDNLL- 78
CC 321 GQAKTKNTVSVNLEITAE-----BWKKKYEKEKNTLRNVIQHLLEML 367
CC 79 -----KAITEKEKIERQSISSPLDNKLVNEDVD-----STKMKLIDDDYDSTKSG 126
CC 368 NRWRNGEAVPEDEQISAKDQK-NLEPCDNTPIIDNAPVAVGISTEEK---EKYDEEISS 423
CC 127 LDHKFQDDPGLHQLDGTPLTAEDIVHKIARIYEEND-----RAVFDKIVSKLNLNGLI 181
CC 424 LYRLQDDKDDIENQ-----QSOLAEKJLQKQMLDDELLASTRDYKEIKQEELTRLQIE 476
CC 182 TESQAHTELEDAVRLQKLSKEANNYE---EDNPKPTSWTENQAGIKPEKVTMAIAD 238
CC 477 NEA-----AKDEVKEVLQAL-BELAVNYDQKSQEVEDKTRANEQUTDELAKQITLTITTOR 531
CC 239 GLAK-----GENDETSTNLTITLNGLERR-----TK 264
CC 532 ELSQQLSBNHKKRATVILNLLKDLGEIGGIITND---VKTADVNGVIEEFTMAR 588
CC 265 TYSEDNFEELQYFPNFALLKS--IDSEKAKEKETLITMTKILDFVKMVKGTISPE 322
CC 589 LYISKMKSEKSVLNRSEKQLESQMSDNRRKNASERELAAQCLLSQHEAKIKSLT--- 644
CC 323 EGVSYLENLDEMIALQTKNKKLEK---NATDNISKLPAPSEKSGHEEDVSTKE----- 371
CC 645 ---DYMGNME-----QKRQLSESDLSBELAKL--RAQEKHVEVSFQKEKHLTRLQ 694
CC 372 EAAKMEKEYGSLKSDTKDDNSNPGGKTDRPKGKTEAYLEAIRKNIEWLKKHKDKGNKEDY 431

Db 695 DAEMKKALEQQMESHRAHQKLSRLDETEEKQKIIDEIRDLNQLQLEQEK-LSSDY 753
QY 432 DLSKMRDPINKQADAYVEKGIL--DKEE 457
Db 754 NKLKIED--QEREMKLEKLLLLNDKRE 778
RESULT 14
ID MLPI YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLPI.
GN MLPI OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Kosling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01992; AAA34783.1; -
CC EMBL; X73541; CAAS1948.1; -
CC EMBL; Z28320; CAA82174.1; -
CC PIR; S38173; S38173.
CC SGD; S0001803; MLPI.
CC GO; GO:0005635; C:nuclear membrane; IDA.
CC GO; GO:0005854; C:nucleoplasm; IDA.
CC GO; GO:0006606; P:protein-nucleus import; IDA.
CC Coiled coil; DNA repair.
CC DOMAIN 69 487 COILED COIL (POTENTIAL).
CC DOMAIN 531 1678 COILED COIL (POTENTIAL).
CC DOMAIN 1834 1866 COILED COIL (POTENTIAL).
CC CONFLICT 301 301 R -> A (IN REF. 1).
CC SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;
CC
CC Query Match 6.9%; Score 166.5; DB 1; Length 1875;
CC Best Local Similarity 23.6%; Pred. No. 1;
CC Matches 112; Conservative 85; Mismatches 178; Indels 99; Gaps 28;
CC
CC 30 SLNRELASRPNEQIAEAEEDKIKKTPPENKPGQSNYSFVDNLLKKAITE-KEKIE 88
CC 588 TLKSKMDLESRIEELQLELEE--LKTSTVFNED-----ASYSNV-----TIKQLTETKRDLE 637

QY 89 KERQSIKSPNDKLVNEDVDSTKRNKLI-----DDYDSTKSGLDHKKQDDPDGLHQLD 143
 Db 638 SQVQD-----LQTRISQITRESTENSLNKEIQDLYDS-KSDISIK-----LGKEKS 684
 QY 144 TPLTAED---IVHKIAARIYEEND--RAVFDKIVSKLNLGLITVESQAHTLEDEVAEVLQ 198
 Db 685 SRILAEERFKLLSNTLDLTKAENDQLRKRFYDLQNTILK-----QDSKTHETLNEVSVCKS 740
 QY 199 KL--ISKANNYEDPNKPTSWTEN---OAGKI-PEK-----VTPMAAIQ-----DGLAK 242
 Db 741 KLSIVETELLNKKERQKRLVHLKQELNKLSPKDSLRIMVTQLTLQKEREDLLE 800
 QY 243 -----GENDETVSNLTITLNGLERRTKYSEDNFEELQYFPNFALLKSIDSEKEA 293
 Db 801 TRKSCQKIDLEALSSELKXKETSQKHIIKQLEEDNNSNTWQNKTEALK----- 852
 QY 294 KEKETLIT-INKTLTDFVKMVMKYGTISPVEGVSVLENLDEMIALQTKNLEKKNATONIS 352
 Db 853 KYESVITSVDKQTDIEKLYKVSLEKE-----IE--EDKIRLHTYVWDETINDSL 905
 QY 353 KLFAPSPKSHETDSTKEAAKMEKYGSLKSDTKDSDNSNPGKGTDPKGTAYLEAI 412
 Db 906 R-----KELEKSKINLTDAYSQIKYKDLVETTSQSLQOTNSKLDL--SFKDFTNQI 955
 QY 413 RKNIEWLKHDKGNKED-YDLSKMRDP-INKQADAYVEKGLDKKEAEAIKRI 464
 Db 956 -KNIT-----DEKTSLEDKISLLKQMFNANLDEL--LQKGMKEKADFKRI 1001

RESULT 15
 CENE_HUMAN
 AC Q02224; STANDARD; PRT; 2663 AA.
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENPE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RA MEDLINE=95196755; PubMed=7889940;
 RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and HUBRI";
 RL J. Cell Biol. 143:49-63(1998).
 CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -!- SUBUNIT: INTERACTS WITH CENP-F AND HUBRI KINASE.
 CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z15005; CAA78727.1; --
 DR PIR; S28261; S28261.
 DR HSP; P17119; 3KAR.
 DR Genew; HGNC:1856; CENPE.
 DR GK; Q02224; --
 DR MIN; 117143; --
 DR GO; GO:0005699; C:kinetochore; TAS.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:008350; F:kinetochore motor activity; TAS.
 DR GO; GO:000067; P:DNA replication and chromosome cycle; TAS.
 DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
 DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN-MOTOR
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP BIND 86 93 ATP (BY SIMILARITY).
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL380C8C8C8 CRC64;
 Query Match 6.98; Score 166; DB 1; Length 2663;
 Best Local Similarity 21.88; Pred. No. 1.7;
 Matches 119; Conservative 103; Mismatches 175; Indels 150; Gaps 28;
 QY 26 SQKSLNRLSALRPLN---EQIAAEEDKIKKTPPENKPGQSNYSFVDNLNLKAIT 82
 Db 1102 AQEK---NHAIKKELSRICDLAEVEEKLKESQQLQKQQL-----LNQVQEMS 1151
 QY 83 EKEKIEKERSIRSPDLNKLNVEDVDSTK---NRKLIDDYDSTKS-----GLDHK 130
 Db 1152 EMQKKINEIENLKNLKNELTLEHMETERLELAQKLNENVEVKSTKERRKVLKELQKS 1211
 QY 131 FQDDPDGL-----HQLDGTPL-TAEDI-----VH-KIAARIYEENDRAVFDKIVSKLNLGL 180
 Db 1212 FETERDHLRGYIREIEATGLQTKELKIAHILKHEQETIDELRRSVSEK-TAQIINTQD 1270
 QY 181 ITESQAHTLEDEVAEVL---QKLIS--KEANNYEEDDNKPTSWTENQAGKIPKVTPTMAA 235
 Db 1271 LEKS--HTKLOEIPVLHBEQELLPNVKVSETQETNWELELLTEQSTTK---DSTTLAR 1325
 QY 236 IQ-----DGLAK-GENDETIVSNLT-----TLTNGLERRTK 264
 Db 1326 IEMERLRLNEKFQESQSEIKSLTKERNLKTKEALEVHDKLKEHRETLAKIQESQSK 1385
 QY 265 TYSEDNPEEL-----QYFPNFYALK-----SISEKE-AKEKE 297
 Db 1386 QEOSLNKMKXDNETTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDMKSVAKED 1445
 QY 298 TLTITMKTLLDFVKMVMKYGTISPVEGVSVLENLDEMIA--LQTKNLE-----KNATD 349
 Db 1446 DLQRLQEVL-----QSSDQLKENIKKIVAKHLETEEBELVAHCLKEQEE 1491
 QY 350 NISKLFPAPSEKSHETDSTKEAAKMEKYGSLKSDTKDSDNSNPGKGTDPKGTKE--- 406
 Db 1492 TINELRVNLSK-----ETEISTIQKLEAINDKLN-----KIQEIVEKEEQLN 1536
 QY 407 -AVLEAIRKNIEWLKHDKKGNKEDVDL-----SKMRDFINKQADAYVEKGLDKKEAEAI 461
 Db 1537 IKQISEVQENVNLEKQPKHRRKAKDSALQSIKSMLELTNRLQESQSEIQIMIKEK-EEM 1595

Qy 462 KRIYSSL 468
||: :|
Db 1596 KRVQEAL 1602

Search completed: January 2, 2004, 11:00:01
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 40 Seconds
(without alignments)
3019.215 Million cell updates/sec

Title: ABU66614

Perfect score: 2399

Sequence: 1 MGFLTGTWLVLPQAF.....EKGILDKKEBAIKRIYSSL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.*
1: sp_archea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2062	86.0	466	11 Q8R1D7	Q8R1D7 mus musculus
2	1216	50.7	457	13 Q91847	Q91847 xenopus lae
3	198	8.3	7210	5 Q9V788	Q9V788 drosophila
4	198	8.3	9270	5 Q8MLD9	Q8MLD9 drosophila
5	197	8.2	2081	10 Q9LH98	Q9LH98 arabidopsis
6	196.5	8.2	1175	16 Q8XNW6	Q8XNW6 clostridium
7	196	8.2	2139	5 Q07569	Q07569 entamoeba h
8	191.5	8.0	2760	5 Q815Y2	Q815Y2 plasmodium
9	189	7.9	1434	5 Q8I492	Q8I492 plasmodium
10	185.5	7.9	1661	5 Q06166	Q06166 plasmodium
11	184	7.7	951	5 Q96229	Q96229 plasmodium
12	183.5	7.6	1790	3 Q07380	Q07380 saccharomyc
13	183.5	7.6	2612	5 Q815X5	Q815X5 plasmodium
14	180.5	7.5	3504	5 Q8IL45	Q8IL45 plasmodium
15	179.5	7.5	1320	11 Q9JK25	Q9JK25 rattus norv
16	179.5	7.5	1510	5 Q25920	Q25920 plasmodium

17	179	7.5	853	10 Q9LFE4	Q9LFE4 arabidopsis
18	179	7.5	1558	5 Q96275	Q96275 plasmodium
19	179	7.5	1786	5 Q9U0P0	Q9U0P0 plasmodium
20	179	7.5	2269	5 Q26223	Q26223 plasmodium
21	179	7.5	2747	5 Q9BJX9	Q9BJX9 plasmodium
22	178.5	7.4	1323	5 Q9NB35	Q9NB35 plasmodium
23	178.5	7.4	3317	16 Q8SWP8	Q8SWP8 mycoplasma
24	178	7.4	840	5 Q815X4	Q815X4 plasmodium
25	177.5	7.4	1464	5 Q81IF6	Q81IF6 plasmodium
26	177	7.4	1112	16 Q8EWQ1	Q8EWQ1 mycoplasma
27	177	7.4	1871	10 Q9SRD5	Q9SRD5 arabidopsis
28	176.5	7.4	682	5 Q8IBV6	Q8IBV6 plasmodium
29	175.5	7.3	1964	5 Q8SWQ7	Q8SWQ7 loligo peal
30	175	7.3	10578	5 Q8ISF5	Q8ISF5 caenorhabdi
31	175	7.3	18519	5 Q8ISF6	Q8ISF6 caenorhabdi
32	175	7.3	18534	5 Q8ISF7	Q8ISF7 caenorhabdi
33	174.5	7.3	924	5 Q15738	Q15738 dictyosteli
34	174.5	7.3	1263	5 Q8IK49	Q8IK49 plasmodium
35	173.5	7.2	1795	5 Q8IE35	Q8IE35 plasmodium
36	172.5	7.2	1134	2 Q99051	Q99051 streptococ
37	172.5	7.2	1979	5 Q96133	Q96133 plasmodium
38	172	7.2	495	5 Q8IEK6	Q8IEK6 plasmodium
39	171.5	7.1	2083	5 Q9N435	Q9N435 caenorhabdi
40	171.5	7.1	3484	5 P91257	P91257 caenorhabdi
41	171.5	7.1	5507	5 Q8IHN3	Q8IHN3 plasmodium
42	170.5	7.1	2771	5 Q26216	Q26216 plasmodium
43	170	7.1	1140	4 Q9ULE5	Q9ULE5 homo sapien
44	170	7.1	1804	11 Q921Q1	Q921Q1 mesocricetu
45	170	7.1	3130	5 Q9BK46	Q9BK46 plasmodium

ALIGNMENTS

RESULT 1

Q8R1D7 ID Q8R1D7 PRELIMINARY; PRT; 466 AA.
AC Q8R1D7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to secretogranin III.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024785; AAH24785.1; -;
SQ SEQUENCE 466 AA; 52783 MW; AB1CB9667276315C CRC64;

Query Match 86.0%; Score 2062; DB 11; Length 466;
Best Local Similarity 87.0%; Pred. No. 2.1e-94;
Matches 410; Conservative 17; Mismatches 36; Indels 8; Gaps 2;

Qy	1	MGFLTGTWLVLPQAFKPGSGQDSKLNHRELSEAPLNQIAEAEEDKIKKT	57	1	MGFLTGTWLVLPQAFKPGSGQDSKLNHRELSEAPLNQIAEAEEDKIKKT	57
Db	1	MGFLTGTWLVLPQAFKPGSGQDSKLNHRELSEAPLNQIAEAEADKIKKA	60	1	MGFLTGTWLVLPQAFKPGSGQDSKLNHRELSEAPLNQIAEAEADKIKKA	60
Qy	58	YPENKPGQSNYSFVNLLNKAITKEKEKERQSRSSPLDNKLNVEDVDSTGNKLI	117	58	YPENKPGQSNYSFVNLLNKAITKEKEKERQSRSSPLDNKLNVEDVDSTGNKLI	117
Db	61	FPSEKPSSESNYSVDNLLNLLRAITKETVEKERQSRSPFDNQLNVEDDSTGNKLI	120	61	FPSEKPSSESNYSVDNLLNLLRAITKETVEKERQSRSPFDNQLNVEDDSTGNKLI	120
Qy	118	DDYDSTKSGLDHFKQDDPDLGLHLDGTPPLPAEDIVHKIAARIYEENDRAVDFDKIVSKLN	177	118	DDYDSTKSGLDHFKQDDPDLGLHLDGTPPLPAEDIVHKIAARIYEENDRAVDFDKIVSKLN	177
Db	121	DEYDSTKSGLDHFKQDDPDLGLHLDGTPPLPAEDIVHKIAARIYEENDRAVDFDKIVSKLN	180	121	DEYDSTKSGLDHFKQDDPDLGLHLDGTPPLPAEDIVHKIAARIYEENDRAVDFDKIVSKLN	180
Qy	178	LGLITESQAHLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIKEKVTMAAIQ	237	178	LGLITESQAHLEDEVAEVLQKLSKEANNYEEDPNKPTSWTENQAGKIKEKVTMAAIQ	237
Db	181	LGLITESQAHLEDEVAEVLQKLSKEANNYEEDLPKPTSRTEENQDGKIKEKVTMAAAVQ	240	181	LGLITESQAHLEDEVAEVLQKLSKEANNYEEDLPKPTSRTEENQDGKIKEKVTMAAAVQ	240

```

238 DGLAKGENDVTSNTLTNLGLRRTKTYSEDNFELOVFPNFIALLKSIDSEKAKE 297
241 DGFTNRDNDVTSNTLTNLGLRRTNPHREDDFELQVFPNFIALLTSIDSEKAKE 300
298 TLITIMKTLLIDFVQWVKYGTISPFGVSYLENLDEMIALQTKNKLKKNATDNISKLP 357
301 TLITIMKTLLIDFVQWVKYGTISPFGVSYLENLDEMIALQTKNKLKKNATDNISKLP 355
358 PSEKSHETDSTKEAAKMEKEVGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNI 417
356 PPEKSEETDSTKEAAKMEKEVGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNI 415
418 WLKXKDKKGNKEDYDLKQWDFINQADAYVEKGLDKKEEAIAKRIYSSL 468
416 WLKXKDKKGNKEDYDLKQWDFINQADAYVEKGLDKKEEAIAKRIYSSL 466

RESULT 2
Q91847 PRELIMINARY; PRT; 457 AA.
AC Q91847;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Secretogranin III.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Int. pituitary;
RX MEDLINE=96217355; PubMed=8632145;
RA Holthuis J.C.M., Martens G.J.M.;
RT "The neuroendocrine proteins secretogranin II and III are regionally
RT conserved and coordinately expressed with proopimelanocortin in
RT xenopus intermediate pituitary.";
RL J. Neurochem. 66:2248-2256(1996).
DR EMBL; X92872; CA63478.1; -
SQ SEQUENCE 457 AA; 51903 MW; B3096512B4B076D7 CRC64;

Query Match 50.7%; Score 1216; DB 13; Length 457;
Best Local Similarity 57.0%; Pred. No. 1.2e-52;
Matches 269; Conservative 63; Mismatches 106; Indels 34; Gaps 10;

Qy 8 TWILVLPIQAPKPGSQKSLHNRSLAERPLNEQIAEAEDEKIKKTYPPENKPGQS 67
Dy 9 TAIIGVFHVQGFNPISIKQDKVHNRSLSEERPLEEQIAEA--DTVSRGTSTENQMLK 66
Qy 68 NYSFVDNLNLKAIKTEKIEKIERQSISSPLDNKLVNEDVDSTKRLIDYDSTKSL 127
Dy 67 NDSFADDLGLFMSGCKDGVKRSSTGSGFSGNRGLTDDADSTKSLKLAIFYDTTSM 126
Qy 128 DHKQDDPDGLHQLDGTPLTAEIVHKIAARIYENDRAVDKIVSKLNLGLITESQAH 187
Dy 127 DYKVEDPDGLHQLD-APTAEIVRKIATRIYENDRGVDFKIVSKLNLGLITESQAY 185
Qy 188 TLEDEAVLQKLSKEANNVEE-----DPNKPTSWTENQAGKPEKVTPTMAAIOG 239
Dy 186 TLEDEAVLQKLSKEANNVEE-----DPNKPTSWTENQAGKPEKVTPTMAAIOG 234
Qy 240 LAKGENDT--VSNLTLTNLGLRRTKTYSEDNFELOVFPNFIALLKSIDSEKAKE 297
Dy 235 LKNKEDRESSTKNEKDSLST-VERNELSPED---DLQYFPNFRNLKSLNENDVKEK 290
Qy 298 TLITIMKTLLIDFVQWVKYGTISPFGVSYLENLDEMIALQTKNKLKKNATDNISKLP 357
Dy 291 TLLITIMKTLLIDFVQWVKYGTITPEEGVNYLENLDDMLAVQKKNLGNFGSHNAL--- 347
358 PSE-KSHETDSTKEAAKMEKEVGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNI 416
Dy 356 PPEKSEETDSTKEAAKMEKEVGLSKDSTKDDNSNPGGKTDEPKGTEAYLEAIRKNI 415

```

```

348 PSDEKNADSDSTKEAAKMEKEVELLDKSTKPEPND--AAESKPKGAETYLEAIRKNI 405
417 EWLKXKDKKGNKEDYDLKQWDFINQADAYVEKGLDKKEEAIAKRIYSSL 468
406 EWLKXKDKKGNKEDYDLKQWDFINQADAYVEKGLDKKEEAIAKRIYSSL 457

RESULT 3
Q9V7G8 PRELIMINARY; PRT; 7210 AA.
AC Q9V7G8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CG18255 protein.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Betens F.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gilek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buzam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

```

[illegible]

DB 5127 DENDKK-QKEDGATNKSQK-AEADVVPEK--ISBEKVAEIK 5164
::||| ||| : | : : || : : :

RESULT 4
QBMLD9
ID QBMLD9 PRELIMINARY; PRT; 9270 AA.
AC QBMLD9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG18255-PA.
GN STRN-MLCK OR G8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxId=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke T.J., Davenport L.B., Davies P.,
de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fostler C., Gabrielian A.S., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kamal B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., MCPHERSON D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupsky M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
FT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RNL [2]
RR SEQUENCE FROM N.A.
RP Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banon J., An H., Baldwin D., Bazelon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorset E., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Gall R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ikegawa C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J.B., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.E.,
 RA Clapp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003808; AAM70936.1; -
 DR FlyBase; FBgn0013988; Strn-Mick.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig c2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00047; ig; 29.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SMO0408; IGC2; 29.
 DR SMART; SMO0406; IGV; 4.
 DR SMART; SMO0220; S_TKc; 1.
 DR SMART; SMO0219; TyrKc; 1.
 DR PROSITE; PS00835; IG_LIKE; 26.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 KW ATP-binding; Immunoglobulin domain; Transferase.
 SQ SEQUENCE 9270 AA; 1036749 MW; 0DD82D040527C2AD CRC64;

 Query Match 8.3%; Score 198; DB 5; Length 9270;
 Best Local Similarity 23.8%; Pred. No. 0.73;
 Matches 110; Conservative 83; Mismatches 180; Indels 90; Gaps 19;

 QY 35 ELSAERPLNEQIAEEDKIKTTPPENKPGQSNYSFVNDNLKATTEKIEKIERQSI 94
 Db 4757 EIVSEKITDEKAQESQKKEVGS---EAKPKAKV-----LEKSGIEEKLDEKKEQ 4806

 QY 95 RSSPLDNKLANVEDVSTKNRKLIDD--YDSTKSLGHDHFDPPDGLHQLDGTPLTAEDIV 152
 Db 4807 TESAIDEKSQAIEVSEIVSEKITDEKAQESQKKEVGS---EAKPKAKVLEKSGIEEKL 4865

 QY 153 HK-----IAARIYENDRAFDKIVSKLLNLGLITESQAHTLEDEVAELVQ-KLISKEANN 207
 Db 4866 NKKEKQTESAIDEKSQAIEVSEIVSEKITDEKAQESQKKEVGS---EAKPKAKVLEKSGIE 4925

 QY 208 YEEDPNKPTSWTENQAKIEKVTIPMAIIOGLAKGENDETIVNTLTNGRLRTTYS 267
 Db 4926 EKLEDEKKEQTE-----SAIDEKFOAEVSEIVSEKIT-----DEKAESR 4967

 QY 268 EDNFEELQYFPNFYAL--KSIDSEK--EAKKET-----LITIMKTLID-- 308
 Db 4968 KEEVKDSEAKPKAKVLEKSGIEEKLDEKKEQTESAIDEKSQAIEVSEIVSEKITDEK 5027

QY 309 ---FVKMMVKYGTISP-----EEGVSYLENDEMIALQTKNLEKNAATDNISKLPAPS 359
 Db 5028 AQESQKKEVKDSEAKPKAKVLEKSGIEEKLDEKKEQTEKTV---AIDTKSQTVVESE 5084
 QY 360 EKSHEETDSTKEEAAARMEKEYGSLKSDSTKDDNSNPGKTDPEPKGTAYLEAIRKNIWL 419
 Db 5085 IIVLEKISEBEAESQKVE-----LKDS-----EAKSKAKVLEKSKTLKEL 5126
 QY 420 KKHDKKGNKEDYDLKMRDFINKQADAVYKGILOKKEEBAIK 462
 Db 5127 DENDKK-QKEDGATNKSQK--AEAADVVPK--ISEEKVAEIK 5164

 RESULT 5
 Q9LH98 PRELIMINARY; PRT; 2081 AA.
 ID Q9LH98
 AC Q9LH98;
 DT 01-OCT-2000 (TREMELrel. 15, Created)
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE Genomic DNA, chromosome 3, BAC clone: T19N8.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones";
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AP002057; BAB03174.1; -
 SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFPF29 CRC64;

 Query Match 8.2%; Score 197; DB 10; Length 2081;
 Best Local Similarity 25.2%; Pred. No. 0.14; Indels 114; Gaps 21;
 Matches 120; Conservative 68; Mismatches 175;

 QY 27 QDKSLHNRLSAERPLNEQIAEED-----KIKTTPPENKPGQSNYS 70
 Db 707 EDKLENKESQTDSDKDDKQVDDKQEAQIYGESKDDKSVKAGKKEKSKENKKTNNEN 766

 QY 71 FVDNL--NLKATTEKIEKIEK--ERQSISSPLNKLNVEDVSTGNRKLIDDYDSTKSG 126
 Db 767 RVRNKEENVQGNKKESEKVEKGEKSKDAKSVETKON--KKLSSTENR-----DEAKERSG 821

 QY 127 LDHGFQDDPDGLHQLDGTPLTAEDIVHKIARIIVEN--DRAVFDKIVSKLLNLGLITE 183
 Db 822 EDNKEDKEESKDYO-----SVKAKEKNGGVDTNVGNKEDSKDLKDDRIVE 868

 QY 184 SQAHTLED-----EVAELVQLKISKE-----ANNYEEDPNKPTSWTENQAKIPEKVTMA 234
 Db 869 VKANKESMKKKEEVQRNDKSSSTKEVRDFANNMDIDVQKSG--ESVKYKDEK----- 921

 QY 235 AIQDGLAKGENDETIVNTLTNGRLRTTYSSEDNFEELQYFPNFVALLKSIDSEKAK 294
 Db 922 --KEG-NKEENKOTI--NTSSKQKQKOKKKKESKN-----SNMKKEED 962

 QY 295 EKETLITIMKTLIDFVQMMVKYGTISP EEGVSYLENDEMIALQTKNLEKNAATDNISKL 354
 Db 963 KKEVNNELKQEDNKKETTKSENSKLKE-----ENKD-----NKEKESDEDSASK- 1008
 QY 355 FPAPSKSHSETOS--TKEEAAKMEKEYGSLKSDSTKDDNSNPG-----GKTDE 400


```

Db 1009 --NREKKYEYKSKTKEAKKKSQDKKREKXDSERKSKKEKERSDLKAKKKEE 1066
Qy 401 PKGRTAYLEAIRKNIEWLKHKDKGKEDYDLSKORDFINKQADAYVEKGILDKKE 457
Db 1067 TREKKES-----ENHKSKKKDKKHEHDNKSMMKKEE--KKEKKKHEESKSRKKE 1115

RESULT 6
QBXNW6 PRELIMINARY; PRT; 1175 AA.
AC Q8XNW6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Probable exonuclease.
GN SBCC OR CPE0216.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_taxid=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RT Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79922.1; -.
DR InterPro; IPR003439; ABC_transporter.
KW Exonuclease; Complete proteome.
SQ SEQUENCE 1175 AA; 136878 MW; 7C6D2366525019C1 CRC64;

Query Match 8.2%; Score 196.5; DB 16; Length 1175;
Best Local Similarity 22.6%; Pred. No. 0.078; Indels 119; Gaps 21;
Matches 116; Conservative 83; Mismatches 196;

Qy 29 KSLHNRLSAERPLNEQIAEAEEDKIKTYPENKPGQSNYSFVDNL-NLLKAITEKE-- 85
Db 256 KELYDKRIEEB-----SLVSSEIEKSFERVEISNKADKIVFINNLEILKEINKEDLK 311
Qy 86 --KIEKQSRSSPLDNKLNVEDVSTKNRKLIDDYDSTKSGLDHKFDQDPDGLHLDG 143
Db 312 FSELNKLLEELINLEENKLFKEPTKKEKLPDLRLKKEKLES--QKRDILFQIKA 369
Qy 144 TPLTAEDIVHKI-----AARIYEENDRAVDFKIVSK-----LNLG 179
Db 370 DGVKLKEACKKIFEDRSKCDTKLNSIEENKRLNEELKEEKEERKEELFVHEFNKINS 429
Qy 180 LITESQAHTLEDEVAEVLQKLI--SKEANNYEEDPNKPTSWTNOAGKIPKVTMPMAAIQ 237
Db 430 LFILNSYSLDKQFNEIKSEBEVELKYYIKNLTEDEKSEK-----DLKVVSLSKIR 482
Qy 238 DGLAK-----GENDETVSNTLTNLGLERTKTVSENFEE-LQYF 277
Db 483 DKLESLLKETPGDNSIILEKQIKLGEYREKLNKYKEIKNSLESLKT--KNFBEKLT 540
Qy 278 PNFYALLASIDSEKAKEKETLITIMKT--LIDFVMMVKYGTISPESGVYLE----- 329
Db 541 ENQKILL-----EKEVRELKDYINKVVEELAHKLRENLEVEGECPCVCGSTHHELNK 595
Qy 330 -NLDE---MIALQTKNKLKNATDNISKJL---FPAPSEK-----SHEETDSTKEAAK 375
Db 596 INLESNEKTLILLESKEKELKELILFESKIEATLEYENKKIIBELNISIEEVEGVEERLK 655
Qy 376 -MEKEYGSLDKSTKDDNSNPGCTDEPKGTEAYLEAIRKNIEWLKHKDKGKEDYDLS 434
Db 656 FLEBEFNTLKDIEEFNLK-----ENLEKLEKLIK--EENKNLE----- 693
Qy 435 KMRDPINKQADAYVEKGLDKHEEAPAIKRIYSSL 468
Db 694 ---NIFNKAERVILCEKIVREKEIASKIKELDKEL 724

```

```

RESULT 7
Q07569 PRELIMINARY; PRT; 2139 AA.
AC Q07569; Q02504;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Myosin heavy chain.
GN MHCA.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_taxid=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=93295430; PubMed=8515774;
RA Raymond-Denise A., Sansonetti P., Guillen N.;
RT "Identification and characterization of a myosin heavy chain gene
RT (mhca) from the human parasitic pathogen Entamoeba histolytica.";
RL Mol. Biochem. Parasitol. 59:123-131(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RA Guillen N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L03534; AAB48065.1; -.
DR HSSP; P08799; 1MND.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SQ SEQUENCE 2139 AA; 245225 MW; C68307341DBS1DD1 CRC64;

Query Match 8.2%; Score 196; DB 5; Length 2139;
Best Local Similarity 21.0%; Pred. No. 0.17;
Matches 107; Conservative 108; Mismatches 191; Indels 104; Gaps 23;

Qy 20 FPKPGSQDKSLHNRLSAERPLNEQIAEAEEDKIKTYPEN-----KPGQSNYSFV 72
Db 859 FEEBKKKKKEIED-----LKKLAE---EIKKREAAENALASATKTGLEAKIQ 906
Qy 73 D-----NLLKAITEKEKIEKQSRSSPLDNKLNVEDV- 109
Db 907 DLEDKISELSKLSAAELDKQELNKLIENTLEEDKEELKTTIDNLKDKSKLKGDELEV 966
Qy 110 -----STKNRKLDDVDSTKSGLDHKFDQDPDGLHLDGTPLTAEEDIVHKIAARIY 160
Db 967 EITELNSQINTLN-ATVDKKDKTTIAEQESIDEKEDKIEITKLKGIKLEEKDDL----- 1020
Qy 161 EENDR-----AVFDKIVSKLNLGLITESQAHTLEDEVAEVLQKLIKSKANNYEEDPNKPT 216
Db 1021 -EQDRAVDSATKDDIAKLNKTIIT-----ECEDAKDEIAKLEQEELEDEE--NKNKDLTNEL 1073
Qy 217 SWTENQAGKIPKVTMPMAAI-----ODGLAKG-ENDETVSNTLTNLGL-LEERTKTVS 267
Db 1074 QQTOLKLGETEKSIAAQAATAKASDRDTLSQLENEKLTITKLTAKADLEKKISGLK 1133
Qy 268 EDNFEELQYFPNFYALLKSID-----SEKAKEKETLITIMKTLDIFV-KMMVKYGTISP 321
Db 1134 QD-YEDDEDDKN-----KIEGLRNAQRKIKELDUDEITKGADVSQLQKQKEEY----- 1181
Qy 322 EEGVSYLENDEMIATQTKNKLKNATDNISKLPAPSEKSSHEETDSTKEAAK-----ME 377
Db 1182 ESQIAKMQEKEKALGNVDVKNK-EKTIKE--KELEIQSLQKLEKDETEVEKEDAEKKKEIE 1238
Qy 378 KEYGSLDKSTKDDNSNPGCTDEPKGTEAYLEAIRKNIEWLKHKDKGKEDYDLS-SKM 436
Db 1239 KEMKALQEE--KENVESSKNSTKDKKKLEEDNLKDTQKDKLDDMTADNEKLEKAKAKDLBAQL 1297

```

QY 437 RDPINKQADAYVEKGLDKEAEAKRIYS 466
 Db 1298 NEVDNHEKAVADAELNKKQSDKELNS 1327

RESULT 8
 Q815Y2 PRELIMINARY; PRT; 2760 AA.
 AC Q815Y2;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN PFL0315C.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian L.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrrell B.,
 RT "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum."
 RL Nature 419:498-511(2002).
 DR EMBL; AE014845; AAN36152.1; --
 KW Hypothetical protein.
 SQ SEQUENCE 2760 AA; 330896 MW; EE9964C4845181AC CRC64;

Query Match 8.0%; Score 191.5; DB 5; Length 2760;
 Best Local Similarity 21.1%; Pred. No. 0.37;
 Matches 121; Conservative 92; Mismatches 207; Indels 153; Gaps 24;

QY 29 KSLNRLSARPLNEQIAEAE-----DKIKTYPPENKPGQSNVGFVD 73
 Db 1452 KNEAKEEETKNDQNVNEKEMDVNSKNREIVQVHNEIKNTNKEEGKKNLLKEK 1511

QY 74 NLN--LLKAITEKIEKHERQ--SIRSSPLDNKLNVEDVDSTKNRLIDDYDSTKGLDH 129
 Db 1512 EINDCLNDYINKQKKEKKNNWAMYGRPIVKRQNNRNINIKNDLKKLYSSKSESGFNDY 1571

QY 130 KF-----QDDPDGLHQLDGTPLTAE--DIVH-----KTAARIYEEN-- 163
 Db 1572 AFYAERFEVITGYNSEFDYLSIDNOAKNEENKNDIHHNNIIKISKWKENIYENSPF 1631

QY 164 ---DRAVFD-----KIVSKLNLGLITESQA----- 186
 Db 1632 HTYGRPIYEKSKNPNYNIKSTHNAILKRRKTLNKSISINFTGNSSNNKIVR 1691

QY 187 -----HTLEDEVAEVLQKISKANNYEEDP-----NKPTSWTENQAGK 225
 Db 1692 TSIGNNTIDYNNSTIKIHKQN--VEDQGYIDLTKRKLTYDALDEINQTOQKNLK 1749

QY 226 IPEKVTPTMAAIQDLAKGENDEVSTNLTITNG--LERRTKYSENPEELQYFNFVAL 283
 Db 1750 ITENIT-----QVQKHGQN---VSNIIKNTGAMLEIKITRK-GNDNNDEQDFSELKAL 1800

QY 284 LKSIDSEKAKEKETITIMKTLIDFVGMVKYGTISPEEGVSYLENDEM-----I 335
 Db 1801 EK-LKKLKEKLTLEETEE--BEKKRLLELNSLKVEEKKKKKKWMEEMKMKI 1856

QY 336 ALQTKNLEK-----NATDNI SKL-----FPAPSEKSHBEETSTKEEAAKM--EKEK 380
 Db 1857 ELQKKKEELKQYRRQOIRKLEBKKEELKLEEMNKLEEQKREIKKMEEKER 1916

QY 381 GSKLDSTKODNSNPGKGTDPKGTBAYL-----EAIKRNIE--WLKHKDKKGNKEDY 431
 Db 1917 EKLKIKKEQKQKEEMKKQEQKKEQIKLMEEMRKFEQKVKVLEBEKKEKELKKL 1976

QY 432 DLSKMRDFINKQADAYVEKGLDKEAEAKRI 464
 Db 1977 EQKKRE-ERKQWEKEEKKEEKKKEEIKKM 2008

RESULT 9
 Q81492 PRELIMINARY; PRT; 1434 AA.
 AC Q81492;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Mature parasite-infected erythrocyte surface antigen (MESA) or
 DE PfEMP2.
 GN MESA.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
 RA Hall N., Bowman S., Churcher C., Quail M., Barrrell B.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
 RN [2] _
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 RA Fellwell T., Gobie A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sultston J.E., Craig A., Newbold C., Barrrell B.G.;
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 DR EMBL; AL929351; CAD51374.1; --
 SQ SEQUENCE 1434 AA; 168287 MW; AB0005P9DC26C989 CRC64;

Query Match 7.9%; Score 189; DB 5; Length 1434;
 Best Local Similarity 21.1%; Pred. No. 0.23; Mismatches 96; Indels 102; Gaps 23;

QY 27 QDKSLNRLSARPLNEQ-----IAEAEEDKIKKTYPPENKPGQSNVFDNLNLKAI 81
 Db 501 KDKVLGEGDKEDVKENDEQKDKVLGEGDKEDVK-----EKNDGKDK-KVIGSEKTOKEI 554

QY 82 TEKEKIEKE-----RQSIKSPDLNKLNVEDVDSTKNRLIDDYDSTKGLDHKFPD--- 133
 Db 555 --KEKVKRKKKKKKKKKKKKIKEN--DTEGNDKVKGPEII--IEEVKEEIKKQVEDGK 608

QY 134 --DPDGLHQLDGTPLTAEIVHKAARIYE--ENDRAVDFKIVSKLNLGLITESQHT 188
 Db 609 ENDETEGNDKVKGPEIITEVKEEIKKQVEGIGKENDTEGNDKVKGP-----EIIITEVKEE 664

QY 189 LEDEVAEVLQ-----KLISKEANNYEEDPNKPTSWTENQAGKIPKVTPTMAAIQDL 240
 Db 665 IKKQVEGIGKENDTESKDLKIGQIITEVKEEIKKENDTENK-----DKVIGQIITEEV 719

QY 241 AKG--ENDETVSNTLITNGLERTKTYSDNEFEEQLYFPNFALLKSIDSEKAKEKET 298

Db	720	KEGIEND-----TENKDKVIGQEIIITE-----EVKKEIEKQEEKNGKEN	755
Qy	299	LITI-----MKTLLIDFYKMMVXGTTISPEBGV--SYLENLD-----EMIALQTKNKLK	345
Db	760	ILEIKDIVIGQEVIIIEEVKKVIKK--KVEGKIKENHTESKDKVIGQEIIVVEEVKEEIEK	816
Qy	346	NATNISKLPAPSEKSHETDSTKEEAANKKEYGSLKSTKDDNSPNPGKTDPEPKGT	405
Db	817	QVEEGIK-----ENDTESKDKVIGQEVING---DVNEEGPENKDKVTKQEKVK-	861
Qy	406	EAYLEAIRKNTIEWLKHKDKGKEDYDLSK--MRDFINKQADAVVEKGILDKEEAAIK	462
Db	862	EVKKEVKKVKVKRVKRNKNRKNONVIGKIMKEDVNEKDTANKDKKEIQEKEKEEVK	920
RESULT 10			
Q06166	166	PRELIMINARY; PRT; 1661 AA.	
ID	Q06166		
AC	Q06166;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Mature PARASITE-infected erythrocyte surface antigen (Antigenic protein, PEMP2).		
DE	Plasmodium falciparum.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RP	SEQUENCE FROM N.A.		
RP	RC STRAIN=PALO ALTO;		
RC	MEDLINE=92158014; PubMed=1741020;		
RA	Coppel R.L.;		
RT	"Repeat structures in a Plasmodium falciparum protein (MESA) that binds human erythrocyte protein 4.1.";		
RT	Mol. Biochem. Parasitol. 50:335-347(1992).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RP	RC STRAIN=PALO ALTO;		
RC	Kun J.F.J., Waller K.L., Coppel R.L.;		
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE OF 797-850 FROM N.A.		
RX	MEDLINE=93122844; PubMed=1478701;		
RA	Saul A., Yeganeh F., Howard R.J.;		
RT	"Conservation of repeating structures in the PEMP2/MESA protein of Plasmodium falciparum.";		
RL	Immunol. Cell Biol. 70:353-355(1992).		
DR	EMBL; AF056936; AAC13303.1; -.		
DR	EMBL; S52458; AAB24869.1; -.		
DR	InterPro; IPR001623; DnaJ_N.		
DR	Pfam; PF00226; DnaJ; 1.		
DR	SMART; SM00271; DnaJ; 1.		
DR	PROSITE; PS50076; DnaJ_2; 1.		
KW	Antigen.		
Qy	22	KPGSQDKSLHNRELSAPRLNEQIAEAEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI	81
Db	895	KDTANKKEIQEKEKEEVEKVEKEEVEKVEKEE-----V	937
Qy	82	TEKEIKERQSIKSSPLDNKLVNEDVDSTKVRKLIDDYDSTKSLDGHKFDODPLHLQ	141
Db	938	KEKEEVEKEEVEKVEKEEVEKVEKEDTSKQKEIQ- EKEKEVEKVEKEDTENKDKV	996
Qy	142	DGTPLETAEDIVHKTAARIYEN-----DRAVFDKIVSKLNLGLITESAHTLEDEVAE	195
Db	997	IGQEIIEEIKVEKVRKVGKNNKNNKNNVIVQEIWMEMLRKDTANKDKVIEQEKKE	1056
Qy	196	VLOKLIKEANNYEEDPKPTSWTENOAGKIPKVTPTMAA1QDGLAKGENDETVSNTLTL	255

```

Db 207 HKVEENKSDHDKVEENKSDHDKIEVVKVEEHEEDREEDKKEKSENKKNKDNEND 266
QY 141 LDCTPLTAEDIVHKIAARIYEENDRAVPKIVSKLLNLGLITESQAHTLEDEVAFLQKL 200
Db 267 EONDEISDEDEVDVDDVEEDKNENDIDDDK-----KETDKTHLEEEENEIIIEKE 315
QY 201 IS-KEANNYEEDPNKPTSW-TENQAGKIPEKVTTPMAAIIQDGLAGKENDETVNTLTITNG 258
Db 316 FSDKKKNGKNKDTKKEKSKDTKEKSKDIEK-----EKSKDKEKESK----- 357
QY 259 LERRTKYSEDNPELOYPNFPYALLKSIDSEKAEKERTLITIMKTILDFVGMVKYGT 318
Db 358 -KDKEKKGKKEKE-----KSKDIEKE-KKDKDIEKESK-DTAKKEKDKD 403
QY 319 ISPEEGVSYLENDEMIALQTKNLEKNATONISKLPAPSEKH-----E 364
Db 404 IEKEKS-----KOMEKLNKQNDKKKODNEKK-----KNDKQDIHDDNDENDMBEIE 452
QY 365 ETDSTKEBAKME-----KEYGLKDSTK-DDNSNPGKTDPEKGTAYLEAI 412
Db 453 ENDEDEDEDEDEKKNKNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN 512
QY 413 RKNIEWLKHDKGKNKEDYDLSKMRDP-----INKQADAYVEKGILD 454
Db 513 NENENGENENENKENEKDKNIKEIENVTNANKENYKINKNSEITITKSNID 564

RESULT 12
Q07380
ID Q07380 PRELIMINARY; PRT; 1790 AA.
AC Q07380; P89892;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical 206.5 kDa protein YDLO58W.
GN US01 OR YDLO58W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecher H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 274106; CAA98621.1; -.
DR EMBL; 274105; CAA98620.1; -.
DR SGD; S0002216; US01.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol_p115_C.
DR InterPro; IPR008953; Usol_p115_C_1.
DR Pfam; PF04871; Usol_p115_C_1.
DR Pfam; PF04869; Usol_p115_head; 1.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;

Query Match
Best Local Similarity 7.6%; Score 183.5; DB 3; Length 1790;
Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;

QY 29 KSLHN--RELSARPLNEQIAEAEDEKIKKTYPPENKPGSNYSFVDNLLKAI-----T 82
Db 965 KSLANNYKQMAE--NESLIKAVEE-----SKNESSQLSNLQNKIDSMQ 1008
QY 83 EKKEIKERQSIRSSPLDNKLNVEDVDSTKNRKL-----DDYSTKSGLDHFKFD--- 133
Db 1009 EKENFQIERGSIENKIEQLKTKTISDLEQTKKEIISKSDSKDEYSQISLLKLETATT 1068
QY 134 -DPDGLHOLDGTPLTAEDIVHKIAARIYEENDRAVPDKIVSKLLNLGLITESQAH----- 187
Db 1069 ANDENVNKISLTKTKEEIEAEALAA--YKNLKNLEUTKLESEKALKEVKEEHLKEEK 1126
QY 188 -TLEDEVAEVLQKLSKEAN-----NYEED--PNKPTSWTENQAGKIPEK 229

```

```

Db 1127 IQLEKATETQQLQNSLRANLESLEKEHEDLAAQLKKYBEQIANKEROYNE-EISQLNDE 1185
QY 230 VTPMAAIIQDGLAGKENDETVNTLTITNGLERRT-----KTYSEDN-- 270
Db 1186 ITSTQENESIKK-KNDELEGEVKAMKSTSEGSNLSKSEIDALNLQIKELKKNETNEA 1244
QY 271 -----FEELOYFPNF-----YALLKSIDSEKAK 294
Db 1245 SLESIKSVSESTVKIKELQDECNFKKEVSELEDKJKASEDKNSKYLELQK-ESEKIKE 1303
QY 295 EKETLITIMKTILDFVGMVK----- 315
Db 1304 ELDAKTELKIQLEKTIKNSKAKESSELSRLKTSSEERKNABEQELKKNKNEIQKNQ 1363
QY 316 -----YGTISPE--EGVSYLENDEMIALQTKNLEKNATONISKLPAPSEK 361
Db 1364 AFEKERKLLNEGSSITIQYSEKINTLE--DELIRLQENELKAKEIDNTRSELEKVSLS 1421
QY 362 SHEETDSTKEBAKMEKEYGSLKDS-TKDDNSNPGKTDPE-----KGTAYLEAIRK 414
Db 1422 NDLLEEEKQNTIKSQDEILSYKDKITRNDEKLLSIEDNRKRDLESKEQLRAAQESKAK 1481
QY 415 NIEWLKHDKGKNKEDYDLSKMRDP-----INKQADAYVEKGILD-KEBAEATKR 463
Db 1482 VEGLKLEEESSKEAELEKSEKMM-KLESTIESNETELKSSMETIRK 1530

RESULT 13
Q815X5
ID Q815X5 PRELIMINARY; PRT; 2612 AA.
AC Q815X5;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL0350C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014845; AAN36159.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2612 AA; 311421 MW; C950B6E466E3D1C5 CRC64;

Query Match
Best Local Similarity 7.6%; Score 183.5; DB 5; Length 2612;
Matches 114; Conservative 96; Mismatches 196; Indels 103; Gaps 23;

QY 27 QDKSLHNRRELSARPP---LNEQIAEAEDEKIKKTYPPENKPGSNYSFVDNLLKAIKE 83
Db 991 REKIDNERNRERKENFLNHTTKKELEEKI-----NNKKEKLEMLK--KE 1036
QY 84 KEKIEKERSQIRSSPLDNKLNVE-----DVDSTKNR--KLIDDYDSTKSG 126
Db 1037 LESLEKEKQIIECEYNNLNQKEEIQERNRNLLIKQELNSRMDRYNELLEDLNKKRKE 1096
QY 127 LD--HKFQDDPG-----LHQDQDGTPLTAEDIVHKIAARIYEENDRAVPDKIVSKLLNL 178

```

```

Db 1097 LEMKMLDDVQBERIKFLNEKNMKEKENEINYMKEELRKERILMIEVEKQKVML 1156
Qy 179 GLI-----TESQHTLEDEVAEVLQKLISKEANNYEDPNKPTSWTENQAGKIPE 228
Db 1157 EDIEKNEKMIKNVEKENELKLBIEKERRNMI-----QNLSEKKEFKLYLEQKYKENFE 1212
Qy 229 KVTMPAALQDGLAK--GENDETIVNTLTNGLERRTKTYTSEDNFEELQ--YFNF----- 280
Db 1213 NE-----KSLAKKFDENEKLQNEI-----GNEKRLKHKERDNFEBQKKIYEEFPRNC 1262
Qy 281 ----YALLKSID--SEKAEKETLITIMKTLDIFVMMVKYGTISPBEGVSYL----- 328
Db 1263 EKYESTQKDYDLEEDKNMKYLLIKEOELENYKNI--YLDIESEKELYVQOEKLN 1320
Qy 329 ---ENL---DEMIALQTN-KLEKNATDNISKL-----FPAPSEKSHETDSTKEEAAKME 377
Db 1321 LEKENLLVEKEQIBELKYNKFNKEKEENDIKIKIINLSQOEDLNKEKENIEKEQIE 1380
Qy 378 KEYSGLKSTKDDNSPGGKTDEPKGTEAVLEAIRKNIEWLKHKDK-KNKEDYDLSKM 436
Db 1381 KRKYDL-DEEESLNNDKAQMEESRKIFDEQLDKIKKNEKELLYDELKNREIHLMEKE 1439
Qy 437 RDFINKQADAVVEKGIIDKEBAEAIKRIYSS 467
Db 1440 KEYNKKLEESNKKESLNSPDEL--KEYSS 1468

RESULT 14
Q8IL45 PRELIMINARY; PRT; 3504 AA.
AC Q8IL45
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pf14.0404.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12369864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow J.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014822; AA037017.1; -
KW Hypothetical protein.
SQ SEQUENCE 3504 AA; 408303 MW; B8454D48D55B84F0 CRC64;

Query Match 7.5%; Score 180.5; DB 5; Length 3504;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 104; Conservative 93; Mismatches 172; Indels 139; Gaps 22;

Qy 32 HN-----RELSAE-----RPLNEQIAEAE-----EDKIKKTYPPENKPGQ----- 66
Db 982 HNLGKNEVSDETWNEYIIPSKSIINEQNTPOYEEVLSKNINNDKNGKRINNNDQLE 1041
Qy 67 ----SNYSFVDNLNLLKAITEKEIERQSIIRSPLDN-----KLVNVEDVDSTKNRKLID 118
Db 1042 SPILSNKKTDDIHI-----IEEKEIKD--PIHNNDLENQNHKKHIQD-----DKLE 1088
Qy 119 DYDSTKSGLDHFKQDDPDGLHQLDGTPLTATDIVHKAARIYENRAVFDKIVSKLNLN 178

```

```

Db 1089 EYSTLPNSKEMK-----NISGDNIENLLSNKKEENHKIHIENIEDTTSK 1134
Qy 179 GLITESQAHTLEDEVAEVLQKLISKEANNYEDPNKPTSWTEN-----QAGKI----- 226
Db 1135 DNEKNMNYIHHDLG--KKDISRENTGDISPNKYLKNKENYGLDDKGIKIYBENN 1191
Qy 227 PEKVTMPAALQDGLAKGENDETIV-----SNTLTLTNGLERRTKTYSEDN 270
Db 1192 EYVTVGSGENKVSQEGNSNDNIMPEYKGNKNSQENSEDNIMTEKNGKNKSSSEENTDDN 1251
Qy 271 FEELQYFPNFPVALLKSIDSEKAEKETLITIMKTLDIFVMMVKYGTISPBEGVSYL 330
Db 1252 -----IMTEKNGKDKSSEENANDNI-----MTEKNGKNKSSQ-----N 1285
Qy 331 LDEMIALQ-----TRNKLEKNATDNISKLPAPSEKSHETDST-----KEEAAK----- 375
Db 1286 IDDNIMTEKNGKNNSQENSEDDNIMTEYKGNKNSQENTDDNIMTEKNGKNKSSSEENTDD 1345
Qy 376 --MEYKVSGLKSTKDDNSPGGKTDEPKGTEAVLEAIRKNIEWLKHKDKKGNKEDYDL 433
Db 1346 NIMTEKNG--KDKSSEENANDNVT--EKNGKNKSSSEENADDNIMTEKNGKNKSSSEENADD 1402
Qy 434 SKWRD--FINKQADAVVEKGIIDKEEAE 459
Db 1403 NIIRENGKNKSSQENSEDDNIMTEYGE 1430

RESULT 15
Q9JK25 PRELIMINARY; PRT; 1320 AA.
AC Q9JK25
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CLIP-170.
GN CYN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RA Galjart N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RX MEDLINE=21185938; PubMed=11290329;
RA Akhmanova A., Hoogenraad C.C., Drabek K., Stepanova T., Dortland B.,
RA Verkerk T., Vermeulen W., Burgering B.M., de Zeeuw C.I., Grosveld F.,
RA Galjar N.;
RT "CLASPs are CLIP-115 and -170 associating proteins involved in the
RT regional regulation of microtubule dynamics in motile fibroblasts."
RL Cell 104:923-935(2001).
DR EMBL; AJ237670; CAB92974.1; -
DR InterPro; IPR000938; CAP-GLY.
DR InterPro; IPR001878; Znf.CCHC.
DR Pfam; PF01302; CAP_GLY_2.
DR SMART; SM00343; Znf.C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS50245; CAP_GLY_2; 2.
SQ SEQUENCE 1320 AA; 148264 MW; 87BE0F0463D20E9D CRC64;

Query Match 7.5%; Score 179.5; DB 11; Length 1320;
Best Local Similarity 22.6%; Pred. No. 0.62;
Matches 134; Conservative 77; Mismatches 211; Indels 171; Gaps 27;

Qy 22 KPGSQDQKSLH-NRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKA 80
Db 525 KPFGVDMSLSLQELISA---LQELVTHDQNEVTSUKDHFGTREMFOKEIKALHA 581
Qy 81 ITEKEIERQSIIRSPLD--NKLNVEDV----- 108

```

